

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 10:57:42 ; Search time 3842.15 Seconds  
(without alignments)  
1824.838 Million cell updates/sec

Title: US-09-394-745-6332  
Perfect score: 425  
Sequence: 1 cggacgcgtgggtgcaattt.....tgtggtgcctctctcaacct 425

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*

28: em\_un:\*  
 29: em\_vi:\*  
 30: em\_htgo\_hum:\*  
 31: em\_htgo\_inv:\*  
 32: em\_htgo\_rod:\*  
 33: em\_htg\_hum:\*  
 34: em\_htg\_inv:\*  
 35: em\_htg\_rod:\*  
 36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

		%					Description
Result	Query	Match	Length	DB	ID		
No.	Score						
	1	35.6	8.4	116841	2	AP000643	AP000643 Homo sapi
c	2	35.6	8.4	153084	2	AC079888	AC079888 Oryza sat
c	3	35.2	8.3	114525	2	AF238279	AF238279 Homo sapi
c	4	34.8	8.2	200000	2	AC004618	AC004618 Homo sapi
c	5	34.8	8.2	200000	2	AC004624	AC004624 Homo sapi
	6	34.4	8.1	165909	2	AC079152	AC079152 Homo sapi
c	7	34.2	8.0	110000	2	LMFLCHR34_05	Continuation (6 of
	8	34.2	8.0	207418	2	AC008676	AC008676 Homo sapi
	9	34.2	8.0	340900	1	SME591791	AL591791 Sinorhizo
c	10	33.8	8.0	20343	5	GGVITIIG	X13607 Chicken vit
c	11	33.8	8.0	194575	2	AC023140	AC023140 Homo sapi
	12	33.6	7.9	141307	8	AC084763	AC084763 Oryza sat
	13	33.6	7.9	146921	8	AP002836	AP002836 Oryza sat
c	14	33.6	7.9	162700	2	AC024105	AC024105 Homo sapi
c	15	33.6	7.9	178141	2	AC074345	AC074345 Homo sapi
	16	33.6	7.9	179714	8	AP002743	AP002743 Oryza sat
c	17	33.4	7.9	138902	9	HSA213H19	AL109749 Human DNA
	18	33.2	7.8	11541	1	AE003960	AE003960 Xylella f
c	19	33	7.8	143411	10	AC009361	AC009361 Mus muscu
c	20	33	7.8	200792	2	AC087540	AC087540 Mus muscu
c	21	33	7.8	205884	2	AC068241	AC068241 Mus muscu
	22	32.6	7.7	87417	2	AC016571	AC016571 Homo sapi
	23	32.6	7.7	151183	9	AC004932	AC004932 Homo sapi
c	24	32.4	7.6	4360	6	AX180877	AX180877 Sequence
	25	32.4	7.6	7988	10	MMU05265	U05265 Mus musculu
c	26	32.4	7.6	148849	9	AL158837	AL158837 Human DNA
c	27	32.4	7.6	165245	2	AL451050	AL451050 Homo sapi
	28	32.2	7.6	4182	8	NEUATPA	M84191 N.crassa mi
	29	32.2	7.6	71414	2	AC087154	AC087154 Mus muscu
	30	32.2	7.6	109047	2	HSDJ19F5	AL078592 Homo sapi
c	31	32.2	7.6	112022	9	HSAJ9611	AJ009611 Homo sapi
c	32	32.2	7.6	143970	2	AL360271	AL360271 Homo sapi
	33	32.2	7.6	155526	2	AC013371	AC013371 Homo sapi
c	34	32.2	7.6	157493	2	AC027068	AC027068 Homo sapi
c	35	32.2	7.6	163337	2	AL445704	AL445704 Homo sapi
c	36	32.2	7.6	177540	9	AC006538	AC006538 Homo sapi
c	37	32.2	7.6	186510	9	HS451B15	Z98050 Human DNA s
c	38	32.2	7.6	220455	2	AC091740	AC091740 Homo sapi

39	32	7.5	100000	9	AP000091	AP000091 Homo sapi
40	32	7.5	100000	9	AP000195	AP000195 Homo sapi
41	32	7.5	131375	2	AC090120	AC090120 Oryza sat
42	32	7.5	146391	8	AC074354	AC074354 Genomic S
43	32	7.5	147771	2	AC021860	AC021860 Homo sapi
c 44	32	7.5	159121	9	AP000236	AP000236 Homo sapi
c 45	32	7.5	161002	9	AL445664	AL445664 Human DNA

# ALIGNMENTS

RESULT 1  
AP000643

LOCUS AP000643 116841 bp DNA HTG 30-MAY-2000

DEFINITION Homo sapiens chromosome 11 clone CMB9-67M21 map 11q22, WORKING DRAFT SEQUENCE, 15 unordered pieces.

ACCESSION AP000643

VERSION AP000643.2 GI:8118835

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE Homo sapiens DNA, clone:CMB9-67M21.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 116841)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Homo sapiens 116,841 genomic DNA of 11q22

JOURNAL Published Only in DataBase (1999) In press

REFERENCE 2 (bases 1 to 116841)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission

JOURNAL Submitted (28-OCT-1999) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923, Fax:81-42-778-9924)

COMMENT On May 31, 2000 this sequence version replaced gi:6997520.

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft11

Center clone name: CMB9-67M21

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 102653 bases at least Q40

Consensus quality: 109653 bases at least Q30

Consensus quality: 113592 bases at least Q20

Insert size: 115441; sum-of-contigs

Quality coverage: 4.10x in Q20 bases; sum-of-contigs

-----  
 NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1	17218 contig of	17218 bp in length
17319	27299 contig of	9981 bp in length
27400	37805 contig of	10406 bp in length
37906	48937 contig of	11032 bp in length
49038	58240 contig of	9203 bp in length
58341	68900 contig of	10560 bp in length
69001	77086 contig of	8086 bp in length
77187	85247 contig of	8061 bp in length
85348	93093 contig of	7746 bp in length
93194	99197 contig of	6004 bp in length
99298	104114 contig of	4817 bp in length
104215	109568 contig of	5354 bp in length
109669	112417 contig of	2749 bp in length
112518	114990 contig of	2473 bp in length
115091	116841 contig of	1751 bp in length

Sequence updated (26-May-2000).

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 15 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

* 1	17218: contig of	17218 bp in length
* 17219	17318: gap of	100 bp
* 17319	27299: contig of	9981 bp in length
* 27300	27399: gap of	100 bp
* 27400	37805: contig of	10406 bp in length
* 37806	37905: gap of	100 bp
* 37906	48937: contig of	11032 bp in length
* 48938	49037: gap of	100 bp
* 49038	58240: contig of	9203 bp in length
* 58241	58340: gap of	100 bp
* 58341	68900: contig of	10560 bp in length
* 68901	69000: gap of	100 bp
* 69001	77086: contig of	8086 bp in length
* 77087	77186: gap of	100 bp
* 77187	85247: contig of	8061 bp in length
* 85248	85347: gap of	100 bp
* 85348	93093: contig of	7746 bp in length
* 93094	93193: gap of	100 bp
* 93194	99197: contig of	6004 bp in length
* 99198	99297: gap of	100 bp
* 99298	104114: contig of	4817 bp in length
* 104115	104214: gap of	100 bp
* 104215	109568: contig of	5354 bp in length
* 109569	109668: gap of	100 bp
* 109669	112417: contig of	2749 bp in length

```

* 112418 112517: gap of      100 bp
* 112518 114990: contig of 2473 bp in length
* 114991 115090: gap of      100 bp
* 115091 116841: contig of 1751 bp in length.

```

```

FEATURES             Location/Qualifiers
    source            1. .116841
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /chromosome="11"
                       /map="11q22"
                       /clone="CMB9-67M21"
    misc_feature      1. .17218
                       /note="assembly_fragment"
    misc_feature      17319. .27299
                       /note="assembly_fragment clone_end:SP6 vector_side:left"
    misc_feature      27400. .37805
                       /note="assembly_fragment"
    misc_feature      37906. .48937
                       /note="assembly_fragment"
    misc_feature      49038. .58240
                       /note="assembly_fragment"
    misc_feature      58341. .68900
                       /note="assembly_fragment clone_end:T7 vector_side:right"
    misc_feature      69001. .77086
                       /note="assembly_fragment"
    misc_feature      77187. .85247
                       /note="assembly_fragment"
    misc_feature      85348. .93093
                       /note="assembly_fragment"
    misc_feature      93194. .99197
                       /note="assembly_fragment"
    misc_feature      99298. .104114
                       /note="assembly_fragment"
    misc_feature      104215. .109568
                       /note="assembly_fragment"
    misc_feature      109669. .112417
                       /note="assembly_fragment"
    misc_feature      112518. .114990
                       /note="assembly_fragment"
    misc_feature      115091. .116841
                       /note="assembly_fragment"

```

```

BASE COUNT   35694 a  22170 c  21077 g  36500 t   1400 others
ORIGIN

```

```

Query Match      8.4%;  Score 35.6;  DB 2;  Length 116841;
Best Local Similarity  60.2%;  Pred. No. 5.3;
Matches  59;  Conservative  0;  Mismatches  39;  Indels  0;  Gaps  0;

```

```

Qy      59  tgtgctctacttctgcctgatggcccttgctcgtagctgctatggctctgtgtcatgtacac 118
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27788  TGTGCTCTACAGATGCCTGCTTCCCTTTCACTTCTCTGCTATGTTATGACACAGCAAGAG 27847

Qy     119  cacctcggcacaagcaggaaggagtggctacaactcgt 156
          |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 27848  GCCCTCATCAGGAGCTAACCAGATTGGCCACCACTCTT 27885

```

RESULT 2  
 AC079888/c  
 LOCUS AC079888 153084 bp DNA HTG 16-MAR-2001  
 DEFINITION Oryza sativa chromosome 10 clone OSJNBa0078001, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 4 unordered pieces.  
 ACCESSION AC079888  
 VERSION AC079888.7 GI:13357270  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;  
 Oryza.  
 REFERENCE 1 (bases 1 to 153084)  
 AUTHORS Buell,R., Hsiao,J., Zismann,V., Moffat,K.M., Hill,J.,  
 Gansberger,K., Burgess,S., Jarrahi,B., Shvartsbeyn,M., Brenner,M.,  
 Cieccko,A., Pai,G., Vanaken,S., Hansen,C., Utterbach,T.,  
 Feldblyum,T., Khalak,H.G., Yuan,Q., Quackenbush,J., White,O.,  
 Salzberg,S. and Fraser,C.  
 TITLE Oryza sativa ssp. japonica cv. Nipponbare OSJNBa0078001 BAC genomic  
 sequence  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 153084)  
 AUTHORS Buell,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-SEP-2000) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA  
 COMMENT On Mar 16, 2001 this sequence version replaced gi:12039424.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 58321: contig of 58321 bp in length  
 \* 58322 58355: gap of unknown length  
 \* 58356 60656: contig of 2301 bp in length  
 \* 60657 60689: gap of unknown length  
 \* 60690 128049: contig of 67360 bp in length  
 \* 128050 128082: gap of unknown length  
 \* 128083 153084: contig of 25002 bp in length.  
 FEATURES Location/Qualifiers  
 source 1. .153084  
 /organism="Oryza sativa"  
 /cultivar="Nipponbare"  
 /sub\_species="japonica"  
 /db\_xref="taxon:4530"  
 /chromosome="10"  
 /clone="OS38-OSJNBa0078001"  
 /clone="OSJNBa0078001"  
 BASE COUNT 44125 a 33372 c 33060 g 42426 t 101 others  
 ORIGIN

Query Match 8.4%; Score 35.6; DB 2; Length 153084;  
Best Local Similarity 47.7%; Pred. No. 5.2;  
Matches 104; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy	23	ggagagagacgagatcatgaggaagcaatactcccctgtgctctacttctgcctgatggc	82
Db	83983	GGCGTGGGTCATCCGCAAGGTGCACCTCGAGTCGCCCCACCTCGCCGTCGGCCTCCTCGG	83924
Qy	83	ccttgctcgtagctgctatggtctgtgtcatgtacaccacctcggcacaagcaggaaggag	142
Db	83923	CCTCGTCGCGTCCTGCCTCGGCACGGTCATGGAGGCGGAGATGGACCGGATCAAACGCAA	83864
Qy	143	tggctacaactcgtaacgaacctgatggaaggggtggatacaactctgttcccatcaacgg	202
Db	83863	GAACGTCGAGCCGTCGCGTCGCTGGCGGCGGCGGCCAGCAACGCTGCCCCGACAACGA	83804
Qy	203	cggtggcgagcccctagctaggcggtggatccgagcctg	240
Db	83803	CGGCGGCGACACCGACCAGATCGAGGACGCCGACGCCG	83766

```

RESULT      3
AF238279/c
LOCUS       AF238279      114525 bp      DNA      HTG      08-JUN-2001
DEFINITION  Homo sapiens chromosome 8 clone RP5-1127D12 map 8p, WORKING DRAFT
SEQUENCE, 26 unordered pieces.
ACCESSION   AF238279
VERSION     AF238279.3  GI:14329033
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 114525)
  AUTHORS   Schilhabel,M.B., Baumgart,C., Blechschmidt,K., Dette,M., Jahn,N.,
              Lehmann,R., Menzel,U., Polley,A., Reichwald,K., Schudy,A.,
              Siddiqui,R., Taudien,S., Wen,G., Siebert,R., Schlegelberger,B.,
              Rosenthal,A. and Platzter,M.
  TITLE     Chromosome 8 genomic sequence
  JOURNAL    Unpublished
REFERENCE   2  (bases 1 to 114525)
  AUTHORS   Genome Sequencing Center Jena.
  TITLE     Direct Submission
  JOURNAL    Submitted (24-FEB-2000) Genome Analysis, Institute of Molecular
              Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Jun 8, 2001 this sequence version replaced gi:8151654.
----- Genome Center
          Center: Insitute of Molecular Biotechnoloy
          Center code: IMB
          Web site: http://genome.imb-jena.de/
          Contact: gscj-submit@genome.imb-jena.de
----- Project Information
          Center project name: H405
          Center clone name: RP5-1127D12
----- Summary Statistics
          Sequencing vector: M13; 100% of reads
          Chemistry: Dye-terminator Big Dye; 100% of reads

```

Assembly program: Phrap; version 0.990329  
Consensus quality: 88803 bases at least Q40  
Consensus quality: 97704 bases at least Q30  
Consensus quality: 104292 bases at least Q20  
Quality coverage: 3.05 x in Q20 bases; sum-of-contigs

-----  
Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality 10. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 26 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\*       1       1583: contig of 1583 bp in length  
\*       1584       1683: gap of unknown length  
\*       1684       2733: contig of 1050 bp in length  
\*       2734       2833: gap of unknown length  
\*       2834       4191: contig of 1358 bp in length  
\*       4192       4291: gap of unknown length  
\*       4292       5780: contig of 1489 bp in length  
\*       5781       5880: gap of unknown length  
\*       5881       6740: contig of 860 bp in length  
\*       6741       6840: gap of unknown length  
\*       6841       7922: contig of 1082 bp in length  
\*       7923       8022: gap of unknown length  
\*       8023       9862: contig of 1840 bp in length  
\*       9863       9962: gap of unknown length  
\*       9963       11563: contig of 1601 bp in length  
\*       11564       11663: gap of unknown length  
\*       11664       12833: contig of 1170 bp in length  
\*       12834       12933: gap of unknown length  
\*       12934       14109: contig of 1176 bp in length  
\*       14110       14209: gap of unknown length  
\*       14210       20044: contig of 5835 bp in length  
\*       20045       20144: gap of unknown length  
\*       20145       23232: contig of 3088 bp in length  
\*       23233       23332: gap of unknown length  
\*       23333       27722: contig of 4390 bp in length  
\*       27723       27822: gap of unknown length  
\*       27823       31485: contig of 3663 bp in length  
\*       31486       31585: gap of unknown length  
\*       31586       35150: contig of 3565 bp in length  
\*       35151       35250: gap of unknown length  
\*       35251       40074: contig of 4824 bp in length  
\*       40075       40174: gap of unknown length  
\*       40175       44444: contig of 4270 bp in length



```

*      44445      44544: gap of unknown length
*      44545      51427: contig of 6883 bp in length
*      51428      51527: gap of unknown length
*      51528      57079: contig of 5552 bp in length
*      57080      57179: gap of unknown length
*      57180      64710: contig of 7531 bp in length
*      64711      64810: gap of unknown length
*      64811      71527: contig of 6717 bp in length
*      71528      71627: gap of unknown length
*      71628      79632: contig of 8005 bp in length
*      79633      79732: gap of unknown length
*      79733      88288: contig of 8556 bp in length
*      88289      88388: gap of unknown length
*      88389      97053: contig of 8665 bp in length
*      97054      97153: gap of unknown length
*      97154      106653: contig of 9500 bp in length
*      106654      106753: gap of unknown length
*      106754      114525: contig of 7772 bp in length.

```

```

FEATURES                      Location/Qualifiers
    source                    1. .114525
                               /organism="Homo sapiens"
                               /db_xref="taxon:9606"
                               /chromosome="8"
                               /map="8p"
                               /clone="RP5-1127D12"

```

```

BASE COUNT      32955 a  23434 c  23679 g  31957 t   2500 others
ORIGIN

```

```

Query Match          8.3%;  Score 35.2;  DB 2;  Length 114525;
Best Local Similarity 50.6%;  Pred. No. 7;
Matches 85;  Conservative 0;  Mismatches 83;  Indels 0;  Gaps 0;

```

```

Qy      100  tggctctgtgtcatgtacaccacctcggcacaagcaggaaggagtggctacaactcgtacg 159
          ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      24677  TGGTCAGACACCTCTGAAACATGGGTGAATTATCAGAGAGGCGTCCCTACAATGATTAAA 24618

Qy      160  aacctgatggaagggtggatacaactctgttcccatcaacggcgggtggcagcccctagc 219
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      24617  CACCAAGGGAAGGCTGCCTTCCCTAGTCCGTGACTGGCACC GGAGTTT TGGGTCCACGGA 24558

Qy      220  taggcggtggatccgagcctgtatcagaaatcgaaataatataagact 267
          || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      24557  TAAAACGTGTCTCTCTGTCTCTACCAGAAAATGAAAGAAATTGAAATT 24510

```

```

RESULT      4
AC004618/c
LOCUS      AC004618      200000 bp      DNA      HTG      04-DEC-1998
DEFINITION      Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 24
                unordered pieces.
ACCESSION      AC004618
VERSION      AC004618.1  GI:3962501
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE      human.
    ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 200000)

AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 200000)

AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.

TITLE Direct Submission

JOURNAL Submitted (27-APR-1998) Department of Genetics, Stanford Human Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA

COMMENT On Dec 4, 1998 this sequence version replaced gi:3927817.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 24 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1111: contig of 1111 bp in length

\* 1112 2912: gap of unknown length

\* 2913 4900: contig of 1988 bp in length

\* 4901 6701: gap of unknown length

\* 6702 8088: contig of 1387 bp in length

\* 8089 9889: gap of unknown length

\* 9890 11145: contig of 1256 bp in length

\* 11146 12946: gap of unknown length

\* 12947 14128: contig of 1182 bp in length

\* 14129 15929: gap of unknown length

\* 15930 18417: contig of 2488 bp in length

\* 18418 20217: gap of unknown length

\* 20218 22278: contig of 2061 bp in length

\* 22279 24078: gap of unknown length

\* 24079 26112: contig of 2034 bp in length

\* 26113 27912: gap of unknown length

\* 27913 31852: contig of 3940 bp in length

\* 31853 33652: gap of unknown length

\* 33653 37450: contig of 3798 bp in length

\* 37451 39250: gap of unknown length

\* 39251 43441: contig of 4191 bp in length

\* 43442 45241: gap of unknown length

\* 45242 51701: contig of 6460 bp in length

\* 51702 53501: gap of unknown length

\* 53502 58600: contig of 5099 bp in length

\* 58601 60400: gap of unknown length

\* 60401 64575: contig of 4175 bp in length

\* 64576 66375: gap of unknown length

\* 66376 72514: contig of 6139 bp in length

\* 72515 74314: gap of unknown length

\* 74315 82110: contig of 7796 bp in length

\* 82111 83910: gap of unknown length

\* 83911 90792: contig of 6882 bp in length

\* 90793 92592: gap of unknown length

\* 92593 99418: contig of 6826 bp in length

\* 99419 101218: gap of unknown length

\* 101219 109648: contig of 8430 bp in length

\* 109649 111448: gap of unknown length

\* 111449 121402: contig of 9954 bp in length  
 \* 121403 123202: gap of unknown length  
 \* 123203 137696: contig of 14494 bp in length  
 \* 137697 139496: gap of unknown length  
 \* 139497 161238: contig of 21742 bp in length  
 \* 161239 163038: gap of unknown length  
 \* 163039 178291: contig of 15253 bp in length  
 \* 178292 180091: gap of unknown length  
 \*\* 180092 200000: contig of 19909 bp in length.

FEATURES Location/Qualifiers  
 source 1. .200000  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"

BASE COUNT 38926 a 40001 c 39609 g 40039 t 41425 others  
 ORIGIN

Query Match 8.2%; Score 34.8; DB 2; Length 200000;  
 Best Local Similarity 46.6%; Pred. No. 8.9;  
 Matches 111; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 61 tgctctactttctgcctgatggcccttgctcgtagctgctatgggtctgtgtcatgtacacca 120  
 ||||| | | ||| || ||| | | | | | ||||| | ||  
 Db 66764 TGCTCTGCATGATGCTGCTGCTTCTTTCCCTTCCTCGTCAATCAGGAGTCATGAAGCCCT 66705  
 Qy 121 cctcggcacaagcaggaaggagtggtacaaactcgtagaacctgatggaaggggtggat 180  
 | | | | | | | | | | | | | | | | | | | |  
 Db 66704 CAGCTCCCCATGTGCTCAGTAGCACATGAACTCACACGCAACATCGAACCTTGGAAGAC 66645  
 Qy 181 acaactctgttcccatcaacggcggtggcagcccctagctaggcggtggatccgagcctg 240  
 | | | | | | | | | | | | | | | | | | | |  
 Db 66644 ATTGAGGTTTTTCAGACCTACAACCTGGTACAGCTTCTAGGCCCAATGCTCCAATGAGCCGC 66585  
 Qy 241 tatcagaaatcgaaataatataagactgtcttcaacggatcacactgccgctccccc 298  
 |||| ||| | | | | | | | | | | | | | | | |  
 Db 66584 CATCACAAAAGGTAGCAAAAAAGCAAACCATTCAGCGGATGCCTCCCCGTAACCAGCA 66527

RESULT 5  
 AC004624/c  
 LOCUS AC004624 200000 bp DNA HTG 03-SEP-1999  
 DEFINITION Homo sapiens chromosome 4, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 2  
 unordered pieces.  
 ACCESSION AC004624  
 VERSION AC004624.6 GI:5706769  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 200000)  
 AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.  
 TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 200000)  
 AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.

TITLE Direct Submission  
 JOURNAL Submitted (28-APR-1998) Department of Genetics, Stanford Human Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA  
 COMMENT On Aug 6, 1999 this sequence version replaced gi:5705982..  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 71732: contig of 71732 bp in length  
 \* 71733 96430: gap of unknown length  
 \* 96431 200000: contig of 103570 bp in length.

FEATURES Location/Qualifiers  
 source 1. .200000  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"

BASE COUNT 44599 a 42806 c 44180 g 43717 t 24698 others  
 ORIGIN

Query Match 8.2%; Score 34.8; DB 2; Length 200000;  
 Best Local Similarity 46.6%; Pred. No. 8.9;  
 Matches 111; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 61 tgctctacttctgcctgatggcccttgctcgtagctgctatggctctgtgtcatgtacacca 120  
 ||||| | | ||| | | || | | ||| | | |||  
 Db 44261 TGCTCTGCATGATGCTGCTTGCTTCTTTCCCTTCCTCGTCAATCAGGAGTCATGAAGCCCT 44202  
 Qy 121 cctcggcacaagcaggaaggagtggtctacaactcgtacgaacctgatggaaggggtggat 180  
 | | | | | | | | | | | | | | | | | | | |  
 Db 44201 CAGCTCCCCATGTGCTCAGTAGCACATGAACTCACACGCAACATCGAACCTTGGAAGAC 44142  
 Qy 181 acaactctgttcccatcaacggcgggtggcagcccctagctaggcgggtggatccgagcctg 240  
 | | | | | | | | | | | | | | | | | | | |  
 Db 44141 ATTGAGGTTTTTCAGACCTACAACCTGGTACAGCTTCTAGGCCCAATGCTCCAATGAGCCGC 44082  
 Qy 241 tatcagaaatcgaaataatataagactgtcttcaacggatcacactgccgctccccc 298  
 |||| ||| | | || | | | |||| |||| | | | || ||  
 Db 44081 CATCACAAAAGGTAGCAAAAAGCAAACCATTCAGCGGATGCCTCCCCGTAACCAGCA 44024

RESULT 6  
 AC079152  
 LOCUS AC079152 165909 bp DNA HTG 20-AUG-2000  
 DEFINITION Homo sapiens chromosome UNK clone RP11-181J6, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 33 unordered pieces.  
 ACCESSION AC079152  
 VERSION AC079152.1 GI:9858435  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 165909)  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 165909)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (20-AUG-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

COMMENT

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Web site:<http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

*	1	1350: contig of 1350 bp in length
*	1351	1450: gap of unknown length
*	1451	3921: contig of 2471 bp in length
*	3922	4021: gap of unknown length
*	4022	6326: contig of 2305 bp in length
*	6327	6426: gap of unknown length
*	6427	9396: contig of 2970 bp in length
*	9397	9496: gap of unknown length
*	9497	12326: contig of 2830 bp in length
*	12327	12426: gap of unknown length
*	12427	14557: contig of 2131 bp in length
*	14558	14657: gap of unknown length
*	14658	17357: contig of 2700 bp in length
*	17358	17457: gap of unknown length
*	17458	19713: contig of 2256 bp in length
*	19714	19813: gap of unknown length
*	19814	23293: contig of 3480 bp in length
*	23294	23393: gap of unknown length
*	23394	27050: contig of 3657 bp in length
*	27051	27150: gap of unknown length
*	27151	30985: contig of 3835 bp in length
*	30986	31085: gap of unknown length
*	31086	35023: contig of 3938 bp in length
*	35024	35123: gap of unknown length
*	35124	38388: contig of 3265 bp in length
*	38389	38488: gap of unknown length
*	38489	42048: contig of 3560 bp in length
*	42049	42148: gap of unknown length
*	42149	46314: contig of 4166 bp in length
*	46315	46414: gap of unknown length
*	46415	50334: contig of 3920 bp in length
*	50335	50434: gap of unknown length
*	50435	57053: contig of 6619 bp in length

```

*      57054      57153: gap of unknown length
*      57154      61080: contig of 3927 bp in length
*      61081      61180: gap of unknown length
*      61181      65562: contig of 4382 bp in length
*      65563      65662: gap of unknown length
*      65663      69259: contig of 3597 bp in length
*      69260      69359: gap of unknown length
*      69360      73828: contig of 4469 bp in length
*      73829      73928: gap of unknown length
*      73929      79818: contig of 5890 bp in length
*      79819      79918: gap of unknown length
*      79919      84908: contig of 4990 bp in length
*      84909      85008: gap of unknown length
*      85009      90835: contig of 5827 bp in length
*      90836      90935: gap of unknown length
*      90936      97723: contig of 6788 bp in length
*      97724      97823: gap of unknown length
*      97824      104663: contig of 6840 bp in length
*     104664      104763: gap of unknown length
*     104764      111094: contig of 6331 bp in length
*     111095      111194: gap of unknown length
*     111195      116818: contig of 5624 bp in length
*     116819      116918: gap of unknown length
*     116919      126871: contig of 9953 bp in length
*     126872      126971: gap of unknown length
*     126972      136900: contig of 9929 bp in length
*     136901      137000: gap of unknown length
*     137001      147497: contig of 10497 bp in length
*     147498      147597: gap of unknown length
*     147598      163433: contig of 15836 bp in length
*     163434      163533: gap of unknown length
*     163534      165909: contig of 2376 bp in length.

```

#### FEATURES

	Location/Qualifiers
source	1. .165909 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="UNK" /clone="RP11-181J6"
misc_feature	1. .1350 /note="assembly_name:Contig10"
misc_feature	1451. .3921 /note="assembly_name:Contig11"
misc_feature	4022. .6326 /note="assembly_name:Contig12"
misc_feature	6427. .9396 /note="assembly_name:Contig13"
misc_feature	9497. .12326 /note="assembly_name:Contig14"
misc_feature	12427. .14557 /note="assembly_name:Contig15"
misc_feature	14658. .17357 /note="assembly_name:Contig16"
misc_feature	17458. .19713 /note="assembly_name:Contig17"
misc_feature	19814. .23293 /note="assembly_name:Contig18 clone_end:SP6"

```

vector_side:left"
misc_feature      23394. .27050
                  /note="assembly_name:Contig19"
misc_feature      27151. .30985
                  /note="assembly_name:Contig20"
misc_feature      31086. .35023
                  /note="assembly_name:Contig21"
misc_feature      35124. .38388
                  /note="assembly_name:Contig22"
misc_feature      38489. .42048
                  /note="assembly_name:Contig23"
misc_feature      42149. .46314
                  /note="assembly_name:Contig24"
misc_feature      46415. .50334
                  /note="assembly_name:Contig25"
misc_feature      50435. .57053
                  /note="assembly_name:Contig26"
misc_feature      57154. .61080
                  /note="assembly_name:Contig27"
misc_feature      61181. .65562
                  /note="assembly_name:Contig28"
misc_feature      65663. .69259
                  /note="assembly_name:Contig29"
misc_feature      69360. .73828
                  /note="assembly_name:Contig30"
misc_feature      73929. .79818
                  /note="assembly_name:Contig31"
misc_feature      79919. .84908
                  /note="assembly_name:Contig32"
misc_feature      85009. .90835
                  /note="assembly_name:Contig33"
misc_feature      90936. .97723
                  /note="assembly_name:Contig34"
misc_feature      97824. .104663
                  /note="assembly_name:Contig35"
misc_feature      104764. .111094
                  /note="assembly_name:Contig36"
misc_feature      111195. .116818
                  /note="assembly_name:Contig37"
misc_feature      116919. .126871
                  /note="assembly_name:Contig38"
misc_feature      126972. .136900
                  /note="assembly_name:Contig39"
misc_feature      137001. .147497
                  /note="assembly_name:Contig40"
misc_feature      147598. .163433
                  /note="assembly_name:Contig41"
misc_feature      163534. .165909
                  /note="assembly_name:Contig9"
BASE COUNT      49077 a 31362 c 32471 g 49718 t 3281 others
ORIGIN

```

```

Query Match      8.1%; Score 34.4; DB 2; Length 165909;
Best Local Similarity 32.2%; Pred. No. 12;
Matches 98; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

```

Qy	98	tatgggtctgtgcatgtacaccacctcggcacaagcaggaaggagtggctacaactcgta	157
Db	46157	TACCGGCGGGCCCCAGGGAAAGCTGACACGGGCAGGCGGGAAGGGAGCACTCGACGGCGGA	46216
Qy	158	cgaacctgatggaagggtggatacaactctgttcccatcaacggcggtggcagccccta	217
Db	46217	CCGATGGGAAAGAAGACGCGCTTGACACGCCAGACGCGGACACGAAGGGAGAAGCACTC	46276
Qy	218	gctaggcggtggatccgagcctgtatcagaaatcgaaataatataagactgtcttcaacg	277
Db	46277	ACTTTTCGATGTTGACGAGGCTCTCCTGGCAGTGGAANNNNNNNNNNNNNNNNNNNNNNN	46336
Qy	278	gatcacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgatt	337
Db	46337	NN	46396
Qy	338	aacggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccc	397
Db	46397	NNNNNNNNNNNNNNNNNNNNCCCTTCTCCCTTCCCCCACCTCCCCACCTCCTCCCTCCC	46456
Qy	398	cctc 401	
Db	46457	TCCC 46460	

Query Match 8.0%; Score 34.2; DB 2; Length 110000;  
Best Local Similarity 50.3%; Pred. No. 14;  
Matches 84; Conservative 0; Mismatches 83; Indels 0; Gaps 0;



Qy	216	tagctaggcggtggatccgagcctgtatcagaaatcgaaataatataagactgtcttcaa	275
Db	12375	TCGCTTCGCGCAGCAGTGC GGCCCTGTAACCGCCGTCGGAGCCATCGCTTCCACCGGTCAA	12316
Qy	276	cggatcacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccga	335
Db	12315	AGCACCGCTCTGCCAGGTGCAGAAGCGCCTCGTGGCGGATGAGTCGCATGCTAGCCCCGA	12256
Qy	336	ttaacggctcacgctaccaggcgctctacgcggatgtgccccctagc	382
Db	12255	TGAGCGCACGACGCTGCCACGCGGAAATGCCGCCGTCGCCACTTGC	12209

\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1010: contig of 1010 bp in length  
\* 1011 1110: gap of unknown length  
\* 1111 4701: contig of 3591 bp in length  
\* 4702 4801: gap of unknown length  
\* 4802 19121: contig of 14320 bp in length  
\* 19122 19221: gap of unknown length  
\* 19222 49181: contig of 29960 bp in length  
\* 49182 49281: gap of unknown length  
\* 49282 85478: contig of 36197 bp in length  
\* 85479 85578: gap of unknown length  
\* 85579 122090: contig of 36512 bp in length  
\* 122091 122190: gap of unknown length  
\* 122191 160393: contig of 38203 bp in length  
\* 160394 160493: gap of unknown length  
\* 160494 207418: contig of 46925 bp in length.

FEATURES Location/Qualifiers  
source 1. .207418  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTB-47B11"  
/clone\_lib="CalTech human BAC library B"  
BASE COUNT 56983 a 45265 c 46330 g 58117 t 723 others  
ORIGIN

Query Match 8.0%; Score 34.2; DB 2; Length 207418;  
Best Local Similarity 56.8%; Pred. No. 14;  
Matches 63; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 18 tttagaggagagagacgagatcatgaggaagcaataactcccctgtgctctacttctgcctg 77  
|| ||||| | | | | | ||||| | | | | |  
Db 25935 TTAGAGGATAAACATTACACCAGGAGGAAGCTGAAATGTCCCTCCCCTGATTTATGGTTA 25994  
  
Qy 78 atggcccttgctgtagctgctatggtctgtgtcatgtacaccacctcggca 128  
||| ||||| ||| |||| || | ||| | |||| || |  
Db 25995 ATGAGCCTTGTGTTAGGTGCTGAGGATACGCTAATGACCTTTACCTGGGAA 26045

RESULT 9  
SME591791  
LOCUS SME591791 340900 bp DNA BCT 16-AUG-2001  
DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 10/12.  
ACCESSION AL591791 AL591688  
VERSION AL591791.1 GI:15075538  
KEYWORDS .  
SOURCE Sinorhizobium meliloti.  
ORGANISM Sinorhizobium meliloti  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
Rhizobiaceae; Sinorhizobium.  
REFERENCE 1 (bases 1 to 340900)  
AUTHORS Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,  
Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,  
Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,

Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ramsperger,U.,  
 Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.  
 TITLE From the Cover: Analysis of the chromosome sequence of the legume  
 symbiont Sinorhizobium meliloti strain 1021  
 JOURNAL Proceedings of the National Academy of Sciences of the United  
 States of America. 98 (17), 9877-9882 (2001)  
 PUBMED 11481430  
 REFERENCE 2 (bases 1 to 340900)  
 AUTHORS Gouzy,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO  
 EU Consortium  
 COMMENT MELILO EU Consortium:  
 Laboratoire de Biologie Moleculaire des Relations  
 Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,  
 France, Laboratoire de Genetique et Developpement UMR6061-CNRS,  
 Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,  
 France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,  
 Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,  
 D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,  
 Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,  
 B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte  
 des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,  
 B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr  
<http://sequence.toulouse.inra.fr/meliloti.html>.  
 FEATURES Location/Qualifiers  
 source 1. .340900  
 /organism="Sinorhizobium meliloti"  
 /strain="1021"  
 /db\_xref="taxon:382"  
 gene 132. .518  
 /gene="SMc01981"  
 CDS 132. .518  
 /gene="SMc01981"  
 /function="small molecule metabolism; energy transfer;  
 electron transport"  
 /note="Product confidence : putative  
 Gene name confidence : hypothetical  
 predicted by Codon\_usage  
 predicted by Homology  
 predicted by FrameD"  
 /codon\_start=1  
 /transl\_table=11  
 /evidence=not\_experimental  
 /product="PUTATIVE CYTOCHROME C TRANSMEMBRANE PROTEIN"  
 /protein\_id="CAC47094.1"  
 /db\_xref="GI:15075539"  
 /translation="MRIALAALSFSALSLCPVAGLAQEGDAEAGATVFKKCATCHVID  
 KDQNKVGPSLQGVIGRTAGTHADFKYSQAMIDAGKGGLVWDDATLAEYLRNPRAKVKG  
 TKMVFPGLKKDEEIANVIAYLKQHPK"  
 gene 656. .1600  
 /gene="coxM OR SMc01982"  
 CDS 656. .1600  
 /gene="coxM OR SMc01982"  
 /EC\_number="1.9.3.1"  
 /function="small molecule metabolism; energy transfer;  
 electron transport"

```

/Note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE ALTERNATIVE CYTOCHROME C OXIDASE
POLYPEPTIDE II TRANSMEMBRANE PROTEIN"
/protein_id="CAC47095.1"
/db_xref="GI:15075540"
/translation="MAVVVILVLLAVGSVLFHLLSPWWWTPIASNWNYIDNTITITFW
ITGIAFTAVVLFMAYCVLRFRRHPRGNTAAAYEPENRRLEGWLATGTTFGVAAMLAPGLF
VWNQFVTVPQDASEVEVIGQWLWSFRLPGADGKLGTTETRDIAPEPTLGVNRDDAAG
QDDIIIEGGELHLPVGKPKVLMLLRSVDVLHDFYVPEFRAKMDMVPGMITYFWLTPTRT
GTFEILCAELCGVGHPPQMRGTVVVDTEEDYQAWLAEQQTFSQLSASSETRAVPEKVCS
GFPSGIATEQGTGASALFKEERECFGPAATTVAASAAQ"
repeat_region complement(1457..1513)
/Note="Sm-5 OR SMC04646
REPEAT SM-5
predicted by Homology"
/evidence=not_experimental
gene 1634..3415
/ gene="coxN OR SMC01983"
CDS 1634..3415
/ gene="coxN OR SMC01983"
/EC_number="1.9.3.1"
/function="small molecule metabolism; energy transfer;
electron transport"
/Note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE ALTERNATIVE CYTOCHROME C OXIDASE
POLYPEPTIDE I TRANSMEMBRANE PROTEIN"
/protein_id="CAC47096.1"
/db_xref="GI:15075541"
/translation="MMVDVRSIGGEALPPPEVEDVELYHPSWWTRYVFSQDAKIIAI
QYSMTAIAIGMVALVLSWLIRLQLGFPGTFELIDAERYYQFITMHGMIMVIYLLTALF
LGGFGNYLIPLMVGARDMVFPYANMLSZYIYLLAVIVLAASFFTPGGPTGAGWTLYPP
QAVLSGTPGGRDWGIIMMLSSLIIFVIGFTMGGLNYVVTVLQGRARGMTLMRLPLTVW
GIFTATVMALLAFPALFVACVMMLFDRLLGTSFFMPAIVEMGEQLQYGGGSPILFQHL
FWFFGHPEVYIVALPAFGIVSDLISTHARKNIFGYRMMVWAIVIIGGLSFIVWAHMY
VSGMNPYFGFFFATTTLIIAVPTAIKVYNWVLTLRGNIHLTLPMLFALAFIVTFVNG
GLTGLFLGNVVVDVPLSDTMFVVAHFHMVMGVAPIMVIFGAIYHWYPKITGRMLNEAM
GQIHFWVTFIGAYAIFFPMHYLGLIGVPRRYHELGEASFVTTTSIAELNAFISVMALLV
GAAQIVFLFNLAWSLRHGREAGGNPWRAATTLEWQTPETPPPHGNWGNELPVVYRWAYD
YSVPGAPEDFIPQNQPTPGRLSHETVS"
gene 3412..4110
/ gene="coxO OR SMC01984"
CDS 3412..4110
/ gene="coxO OR SMC01984"

```

```

/EC_number="1.9.3.-"
/function="small molecule metabolism; energy transfer;
electron transport"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE CYTOCHROME-C OXIDASE TRANSMEMBRANE
PROTEIN"
/protein_id="CAC47097.1"
/db_xref="GI:15075542"
/translation="MSIVAFFLAAIAAIIAWWLAGQRLTSRPWLEVGHFHDRRGATRL
PPAKIGLGVFLAVVGALFSLAISAYFMRMASSDWGALPLPGLLWLNTGILAAGSITLH
WTKVEAERRNDEAARIGLLAGLALGLAFLAGQLFAWRALSDAGYFLAGNPANSFFYLL
TGMHGLHIIIGGLFALGRVTAHASQTPLGNRRLSIELCAIYWHFMLIVWLVLFLALFAG
WASGVIEFCRQLLT"
gene 4123. .4845
/ gene="coxP OR SMc01985"
CDS 4123. .4845
/ gene="coxP OR SMc01985"
/EC_number="1.9.3.-"
/function="small molecule metabolism; energy transfer;
electron transport"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE CYTOCHROME-C OXIDASE PROTEIN"
/protein_id="CAC47098.1"
/db_xref="GI:15075543"
/translation="MSAPMKKPAAETLPRPAGLAGLASDWASDQRTFKDVSWGKAMMW
IFLLSDTFVFGCFLLAYMSARMSTSVWPWNPSEVFALEIGGTHMPLILIAIMTFVLIS
SSGTMAMAVNYGYRRDRRKTAALMLLTALFGAAFVGMQAFEWKSLIAEGVRPWGNPWG
AAQFGSTFFMITGFHGHVTFGVIFLLIVARKVWRGDFETERRGFFTSRKGRYEIVEI
TGLYWHFVDLVVVFIFAFFYLW"
gene 4857. .5216
/ gene="SMc01986"
CDS 4857. .5216.
/ gene="SMc01986"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by FrameD"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAC47099.1"

```

```

/db_xref="GI:15075544"
/translation="MAHAETHPAQHAAHTEHQHAIKLYLLVWGLLFVLSALSYLVD
YFGLQGYLRWSLILIFMMLKAGLIVAVFMHMAWERLALIYAIILPPLLVLVLFVALMVS
ESNYVLFTRLTFFGGGE"
gene      5477. .8038
          /gene="SMc01987"
CDS       5477. .8038
          /gene="SMc01987"
          /function="small molecule metabolism"
          /note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by FrameD"
          /codon_start=1
          /transl_table=11

```

```

Query Match      8.0%;  Score 34.2;  DB 1;  Length 340900;
Best Local Similarity  50.9%;  Pred. No. 13;
Matches  81;  Conservative  0;  Mismatches  78;  Indels  0;  Gaps  0;

```

```

Qy      1  cggacgcgtgggtgcaatttgaggagagagacgagatcatgaggaagcaatactcccctg  60
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187045 CGGGCAGGCCGTTGGGATTGATGAAATGCGACGAGGTCATGCCGATGCGGCGCGCCTCGG 187104

Qy      61  tgctctacttctgcctgatggcccttgctcgtagctgctatggctctgtgtcatgtacacca 120
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187105 CGTTCATCTTCTCGATGAAGGCCTGTTCCGATCCGCCGACCGTCTCGGCGATCGCGACGG 187164

Qy     121  cctcggcacaagcaggaaggagtggctacaactcgtacg 159
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187165 CGACGTCGTTCCGCCACTTGATCAGCATCATCTTCAACG 187203

```

```

RESULT 10
GGVITIIG/c
LOCUS      GGVITIIG      20343 bp      DNA                      VRT      10-FEB-1999
DEFINITION Chicken vitellogenin II gene.
ACCESSION  X13607
VERSION    X13607.1  GI:63886
KEYWORDS   vitellogenin.
SOURCE     chicken.
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE  1  (bases 1 to 20343)
AUTHORS    Ab,G.
TITLE      Direct Submission
JOURNAL    Submitted (28-NOV-1988) AB G., Dept. of Biochemistry, Nijenborgh
            16, 9747 AG Groningen, The Netherlands
REFERENCE  2  (bases 1 to 20343)
AUTHORS    van het Schip,F.D., Samallo,J., Broos,J., Ophuis,J., Mojet,M.,
            Gruber,M. and AB,G.
TITLE      Nucleotide sequence of a chicken vitellogenin gene and derived
            amino acid sequence of the encoded yolk precursor protein
JOURNAL    J. Mol. Biol. 196 (2), 245-260 (1987)

```

```

MEDLINE      88011328
COMMENT      The sequence overlaps with that reported by Byrne et. al. in
              Biochemistry 23:4275-4279(1984) K02113, by Nardelli et. al. in J.
              Biol. Chem. 262:15377-15385(1987) M18060, by Nardelli et. al.
              Biochemistry 26:6397-6402(1987) and Walker et. al. in EMBO J.
              2:2271-2279(1983).
FEATURES
source       Location/Qualifiers
              1. .20343
              /organism="Gallus gallus"
              /strain="White Leghorn"
              /db_xref="taxon:9031"
              /tissue_type="blood"
              /clone_lib="Charon 4A"
              /clone="22, 24"
exon         1. .53
              /number=1
CDS          join(14. .53,169. .189,290. .441,1769. .2020,2532. .2693,
              2799. .2947,3116. .3266,3385. .3542,3625. .3790,4692.
              .4817,
              5222. .5440,6011. .6228,6345. .6442,6920. .7044,7166.
              .7267,
              7906. .8025,8112. .8301,9055. .9148,9281. .9466,9959.
              .10103,
              10225. .10447,10538. .10713,11067. .11756,12221. .12271,
              12525. .12614,13054. .13266,13362. .13504,13889. .14027,
              14568. .14703,15161. .15347,15679. .15770,16857. .16954,
              17704. .17863,18919. .19080,20003. .20121)
              /codon_start=1
              /product="vitellogenin"
              /protein_id="CAA31942.1"
              /db_xref="GI:63887"
              /db_xref="SWISS-PROT:P02845"
              /translation="MRGIILALVLTLVGSQKFDIDPGFNRRRSYLYNYEGSMLNGLQD
              RSLGKAGVRLSSKLEISGLPENAYLLKVRSPQVEEYNGVWPRDPFTRSSKITQVISSC
              FTRLFKFEYSSGRIGNIYAPEDCPDLCVNIVRGILNMFQMTIKKSQNVYELQEAGIGG
              ICHARYVIQEDRKNRIYVTRTVDLNNCQEKVQKSIGMAYIYPCVPDVMKERLTKGTT
              AFSYKCLKQSDSGTLITDVSSRQVYQISPFNEPTGVAVMEARQQLTLVEVRSEGRSAPD
              VPMQNYGSLRYRFPVAVLPQMPLQLIKTNPEQRIVETLQHIVLNNQQDFHDDVSYRFL
              EVVQLCRIANADNLESIWRQVSDKPRYRRWLLSAVSASGTTETLTKFLKNRIRNDLNY
              IQTLLTVSLTLHLQADEHTLPIAADLMTSSRIQKNPVLQQVACLGYSVVNRYCSQT
              SACPKEALQPIHDLADEAISRGREDKMKLALKCIGNMGEPASLKRILKFLPISSSSAA
              DIPVHIQIDAITALKKIAWKDPKTVQGYLIQILADQSLPPEVRMMACAVIFETRPALA
              LITTIANVAMKESNMQVASFVYSHMKSLSKSRLPFMYNISSACNIALKLLSPKLDMS
              YRYSKVIADTYFDNYRVGATGEIFVNSPRTMFPSAIISKLMANSAGSVADLVEVGI
              RVEGLADVIMKRNIPFAEYPTYKQIKELGKALQGWEKLPPTETPLVSAYLKILGQEVAF
              ININKELLQQVMKTVVEPADRNAAIKRIANQIRNSIAGQWTQPVWMGELRYVVPSCLG
              LPLEYGSYTTALARAASVEGKMTPLTGDFRLSQLESTMQIRSDLKPSLYVHTVAT
              MGVNTEYFQHAVEIQGEVQTRMPMKFIDAKIDVKLKNLKIETNPCEETEIVVGRHKAF
              AVSRNIGELGVEKRTSILPEDAPLDVTEEPFQTSEASREHFAMQGPDSMPRKQSHSS
              REDLRRSTGKRAHKRDICLKMHHIGCQLCFSSRRSRDASFIQNTYLHKLIGEHEAKIVL
              MPVHTDADIDKIQLEIQAGSRAAARIITEVNPESSEEEDESSPYEDIQAKLKRILGIDS
              MFKVANKTRHPKNRPSKKGNTVLAIEFGTEPDAKTSSSSSSSASSTATSSSSSSASSPNR
              KKPMDDEENDQVKQARNKDASSSSRSKSSNSKRSKSSKSSNSKRSKSSSSSSSSSSSS
              RSSSSSSSSSSNSKSSSSSSKSSSSSSSRSSSSKSSSSSSSSSSSSSSSSSSSSSSSS
              SSKSSSSSHSHSHSGHLNGSSSSSSSSSRSSSVSHSHSGHLEDDSSSSSSSVLSKI
              WGRHEIYQYRFRSAHRQEFPKRKLPGDRATSRYSSTRSSHDTSRAASWPKFLGDIKTP
              VLA AFLHGISNNKKTGGLOLVVYADTDSVRPRVOVFVTNLTDSSKWKL CADASVRNAA

```

KAVAYVKWGWDCRDYKVSTELVTGRFAGHPAAQVKLEWPKVPSNVRSVVEWFYEFVPG  
AAFMLGFSEMRDKNPSRQARMVVALTSPRTCDVVVKLPDIILYQKAVRLPLSLPVGPR  
IPASELQPPiWNVFAEAPSAVLENLKARCSVSYNKIKTfNEVKFNYSMPANCYHILVQ  
DCSSELKFLVMMKSAGEATNLKAINIKIGSHEIDMHPVNGQVKLLVDGAESPTANISL  
ISAGASLWIHNENQGFALAAPGHGIDKLYFDGKTITIQVPLWMAGKTCGICGKYDAEC  
EQEYRMPNGYLAKNAVSFGHSWILEEAPCRGACKLHRSFVKLEKTVQLAGVDSKCYST  
EPVLRCAKGCSATKTTPTVTVGFHCLPADSANS�TDKQMKYDQKSEDMQDTVDAHTTCS  
CENEECST"

intron	54. .168
	/number=1
exon	169. .189
	/number=2
intron	190. .289
	/number=2
exon	290. .441
	/number=3
intron	442. .1768
	/number=3
exon	1769. .2020
	/number=4
intron	2021. .2531
	/number=4
exon	2532. .2693
	/number=5
intron	2694. .2798
	/number=5
exon	2799. .2947
	/number=6
intron	2948. .3115
	/number=6
exon	3116. .3266
	/number=7
intron	3267. .3384
	/number=7
exon	3385. .3542
	/number=8
intron	3543. .3624
	/number=8
exon	3625. .3790
	/number=9
intron	3791. .4691
	/number=9
exon	4692. .4817
	/number=10
intron	4818. .5221
	/number=10
exon	5222. .5440
	/number=11
intron	5441. .6010
	/number=11
exon	6011. .6228
	/number=12
intron	6229. .6344
	/number=12
exon	6345. .6442
	/number=13
intron	6443. .6919



	/number=13
exon	6920. .7044
	/number=14
intron	7045. .7165
	/number=14
exon	7166. .7267
	/number=15
intron	7268. .7905
	/number=15
exon	7906. .8025
	/number=16
intron	8026. .8111
	/number=16
exon	8112. .8301
	/number=17
intron	8302. .9054
	/number=17
exon	9055. .9148
	/number=18
intron	9149. .9280
	/number=18
exon	9281. .9466
	/number=19
intron	9467. .9958
	/number=19
exon	9959. .10103
	/number=20
intron	10104. .10224
	/number=20
exon	10225. .10447
	/number=21
intron	10448. .10537
	/number=21
exon	10538. .10713
	/number=22
intron	10714. .11066
	/number=22
exon	11067. .11756
	/number=23
intron	11757. .12220
	/number=23
exon	12221. .12271
	/number=24
intron	12272. .12524
	/number=24
exon	12525. .12614
	/number=25
intron	12615. .13053
	/number=25
exon	13054. .13266
	/number=26
intron	13267. .13361
	/number=26
exon	13362. .13504
	/number=27
intron	13505. .13888
	/number=27

exon 13889. .14027  
 /number=28  
 intron 14028. .14567  
 /number=28  
 exon 14568. .14703  
 /number=29  
 intron 14704. .15160  
 /number=29  
 exon 15161. .15347  
 /number=30  
 intron 15348. .15678

Query Match 8.0%; Score 33.8; DB 5; Length 20343;  
 Best Local Similarity 51.7%; Pred. No. 21;  
 Matches 77; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 207 ggcagcccctagctaggcggtggatccgagcctgtatcagaaatcgaaataatataagac 266  
 || | | | | | | | | | | | | | | | | | | | | | | | | | |  
 Db 11049 GGAAATCAGAGGAGAGGGGATACATCTGTTTCTGTATTTACAATGTGAAATATCTGTGGG 10990  
  
 Qy 267 tgtottcaacggatcacactgccgctccccacgctaaatttgggggctacagtgcacac 326  
 || | | | | | | | | | | | | | | | | | | | | | | | | | |  
 Db 10989 TGTCAGACAAGCTGCTCACC GCAGCCTTTCCTGCCTCACCATGGAAGCTGCAGAGCTCAC 10930  
  
 Qy 327 gctagccgattaacggctcacgctaccag 355  
 || || | | | | | | | | | | | | | | | | | | | | | | | | |  
 Db 10929 TCTTCCCCACCAAAATCTGCCCCAGCAAG 10901

RESULT 11  
 AC023140/c  
 LOCUS AC023140 194575 bp DNA HTG 07-JUL-2000  
 DEFINITION Homo sapiens chromosome 15 clone RP11-535P8, WORKING DRAFT  
 SEQUENCE, 26 unordered pieces.  
 ACCESSION AC023140  
 VERSION AC023140.4 GI:8570285  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 194575)  
 AUTHORS Waterston,R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 194575)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-FEB-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 COMMENT On Jun 17, 2000 this sequence version replaced gi:7232193.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----

Center project name: H\_NH0535P08

----- Summary Statistics -----

Sequencing vector: M13; 100%

Sequencing vector: plasmid; 0%

Chemistry: Dye-primer ET; 100% of reads

Chemistry: Dye-terminator Big Dye; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 183084 bases at least Q40

Consensus quality: 186527 bases at least Q30

Consensus quality: 188372 bases at least Q20

Insert size: 209000; agarose-fp

Insert size: 192075; sum-of-contigs

Quality coverage: 3.93 in Q20 bases; agarose-fp

Quality coverage: 4.32 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 26 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

* 1	1295: contig of 1295 bp in length
* 1296	1395: gap of unknown length
* 1396	3355: contig of 1960 bp in length
* 3356	3455: gap of unknown length
* 3456	5113: contig of 1658 bp in length
* 5114	5213: gap of unknown length
* 5214	6910: contig of 1697 bp in length
* 6911	7010: gap of unknown length
* 7011	9042: contig of 2032 bp in length
* 9043	9142: gap of unknown length
* 9143	11645: contig of 2503 bp in length
* 11646	11745: gap of unknown length
* 11746	14544: contig of 2799 bp in length
* 14545	14644: gap of unknown length
* 14645	17322: contig of 2678 bp in length
* 17323	17422: gap of unknown length
* 17423	19837: contig of 2415 bp in length
* 19838	19937: gap of unknown length
* 19938	25695: contig of 5758 bp in length
* 25696	25795: gap of unknown length
* 25796	29231: contig of 3436 bp in length
* 29232	29331: gap of unknown length
* 29332	32892: contig of 3561 bp in length
* 32893	32992: gap of unknown length
* 32993	38263: contig of 5271 bp in length
* 38264	38363: gap of unknown length
* 38364	42404: contig of 4041 bp in length
* 42405	42504: gap of unknown length
* 42505	48404: contig of 5900 bp in length
* 48405	48504: gap of unknown length
* 48505	57290: contig of 8786 bp in length
* 57291	57390: gap of unknown length
* 57391	65905: contig of 8515 bp in length

```

*      65906      66005: gap of unknown length
*      66006      74728: contig of 8723 bp in length
*      74729      74828: gap of unknown length
*      74829      87289: contig of 12461 bp in length
*      87290      87389: gap of unknown length
*      87390      99467: contig of 12078 bp in length
*      99468      99567: gap of unknown length
*      99568     111113: contig of 11546 bp in length
*     111114     111213: gap of unknown length
*     111214     122406: contig of 11193 bp in length
*     122407     122506: gap of unknown length
*     122507     135480: contig of 12974 bp in length
*     135481     135580: gap of unknown length
*     135581     153618: contig of 18038 bp in length
*     153619     153718: gap of unknown length
*     153719     170017: contig of 16299 bp in length
*     170018     170117: gap of unknown length
*     170118     194575: contig of 24458 bp in length.

```

```

FEATURES              Location/Qualifiers
    source              1. .194575
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="15"
                        /clone="RP11-535P8"
    misc_feature        1. .1295
                        /note="assembly_name:Contig9"
    misc_feature        1396. .3355
                        /note="assembly_name:Contig10"
    misc_feature        3456. .5113
                        /note="assembly_name:Contig11"
    misc_feature        5214. .6910
                        /note="assembly_name:Contig12"
    misc_feature        7011. .9042
                        /note="assembly_name:Contig13"
    misc_feature        9143. .11645
                        /note="assembly_name:Contig14"
    misc_feature        11746. .14544
                        /note="assembly_name:Contig15"
    misc_feature        14645. .17322
                        /note="assembly_name:Contig16"
                        clone_end:T7
                        vector_side:left"
    misc_feature        17423. .19837
                        /note="assembly_name:Contig17"
    misc_feature        19938. .25695
                        /note="assembly_name:Contig18"
    misc_feature        25796. .29231
                        /note="assembly_name:Contig19"
    misc_feature        29332. .32892
                        /note="assembly_name:Contig20"
    misc_feature        32993. .38263
                        /note="assembly_name:Contig21"
    misc_feature        38364. .42404
                        /note="assembly_name:Contig22"
    misc_feature        42505. .48404
                        /note="assembly_name:Contig23"
    misc_feature        48505. .57290

```

misc\_feature /note="assembly\_name:Contig24"  
 57391. .65905  
 misc\_feature /note="assembly\_name:Contig25"  
 66006. .74728  
 misc\_feature /note="assembly\_name:Contig26"  
 74829. .87289  
 misc\_feature /note="assembly\_name:Contig27"  
 87390. .99467  
 misc\_feature /note="assembly\_name:Contig28"  
 99568. .111113  
 misc\_feature /note="assembly\_name:Contig29"  
 111214. .122406  
 misc\_feature /note="assembly\_name:Contig30"  
 122507. .135480  
 misc\_feature /note="assembly\_name:Contig31"  
 135581. .153618  
 misc\_feature /note="assembly\_name:Contig32"  
 153719. .170017  
 misc\_feature /note="assembly\_name:Contig33"  
 clone\_end:SP6  
 vector\_side:right"  
 misc\_feature 170118. .194575  
 /note="assembly\_name:Contig34"

BASE COUNT 58990 a 40026 c 37747 g 55237 t 2575 others  
 ORIGIN

Query Match 8.0%; Score 33.8; DB 2; Length 194575;  
 Best Local Similarity 53.4%; Pred. No. 18;  
 Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 293 ccccccacgctaaatttgggggctacagtgcacacgctagccgattaacgggctcacgctac 352  
 || | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 Db 71784 CCACTGTGATTTCTTTGCTGCTGACTGGGCCCACCTTTGAAAATCTATGGCTGACACTTC 71725  
 Qy 353 caggcgctctacgcggatgtgccccctagccagcttctctctccccctcggttctgtggtg 412  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 Db 71724 CTTACTATCTAAGATGACATCAGTGCCATCCTGTCTTTCACTCTTCCTCCTACAGGGCTT 71665  
 Qy 413 cctctctcaacct 425  
 || |||| | |  
 Db 71664 TCTATCTCATTCT 71652

RESULT 12  
 AC084763  
 LOCUS AC084763 141307 bp DNA PLN 30-JAN-2001  
 DEFINITION Oryza sativa chromosome 10 BAC OSJNBa0027P10 genomic sequence,  
 complete sequence.  
 ACCESSION AC084763  
 VERSION AC084763.4 GI:12597872  
 KEYWORDS HTG.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;  
 Oryza.

REFERENCE 1 (bases 1 to 141307)  
AUTHORS Buell,C.R., Yuan,Q., Moffat,K.S., Hill,J.N., Jenkins,C.N.,  
Hsiao,J., Zismann,V., Pai,G., Bowman,C.L., Fujii,C.Y.,  
VanAken,S.E., Craven,B., Khalak,H., Feldblyum,T.V., Quackenbush,J.,  
White,O., Salzberg,S.L. and Fraser,C.M.  
TITLE Oryza sativa chromosome 10 BAC OSJNBa0027P10 genomic sequence  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 141307)  
AUTHORS Buell,R.  
TITLE Direct Submission  
JOURNAL Submitted (15-NOV-2000) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
REFERENCE 3 (bases 1 to 141307)  
AUTHORS Buell,R.  
TITLE Direct Submission  
JOURNAL Submitted (30-JAN-2001) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org  
COMMENT On Jan 30, 2001 this sequence version replaced gi:12039410.  
Address all correspondence to:rice@tigr.org

BAC clone OSJNBa0027P10 is from Oryza sativa chromosome 10  
The orientation of the sequence is from SP6 to T7 end of the BAC  
clone.  
Genes were identified by a combination of several methods: Gene  
prediction programs including Genscan and Genscan+ (Chris Burge,  
<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkHMM (Mark Borodovsky,  
<http://genemark.biology.gatech.edu/GeneMark/>), Fgenesh  
(<http://www.softberry.com/>), and GeneSplicer (Mihaela Pertea and  
Steven Salzberg, contact mpertea@tigr.org), searches of the  
complete sequence against a peptide database and the plant EST  
database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated  
genes are named to indicate the level of evidence for their  
annotation. Genes with similarity to other proteins are named after  
the database hits. Genes without significant peptide similarity but  
with EST similarity are named as unknown proteins. Genes without  
protein or EST similarity, that are predicted by more than two gene  
prediction programs over most of their length are annotated as  
hypothetical proteins. Genes encoding tRNAs are predicted by  
tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).  
Simple repeats are identified by repeatmasker (Arian Smit,  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES Location/Qualifiers  
source 1. .141307  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/sub\_species="japonica"  
/db\_xref="taxon:4530"  
/chromosome="10"  
/map="near C239"  
/clone="OSJNBa0027P10"  
repeat\_region complement(784. .934)  
/rpt\_family="Explorer\_Os3 MITE element from gb:D25363 Rice  
genomic DNA, G1103A, sequence tagged site (66 to 225) 160  
nt"  
repeat\_region complement(2007. .2034)  
/rpt\_family="AT\_rich"  
repeat\_region 3436. .3458

```

repeat_region      /rpt_family="(TTAA)n"
                    3916. .4092
                    /rpt_family="Wanderer_Ol1 MITE element from gb:U34601
                    Oryza sativa wanderer mobile element linked to Xa21 (411
                    to 637) 227 nt"
repeat_region      complement(3975. .4073)
                    /rpt_family="Tourist_Ol2 MITE element from gb:U72727 Oryza
                    longistaminata receptor kinase-like protein, family member
                    A2, pseudogene sequence(5453 to 5694) 242 nt"
repeat_region      complement(4684. .4705)
                    /rpt_family="(TA)n"
repeat_region      complement(5218. .5247)
                    /rpt_family="AT_rich"
repeat_region      complement(5736. .5768)
                    /rpt_family="AT_rich"
repeat_region      complement(5950. .6118)
                    /rpt_family="Gaijin_Os3 element from gb:D32165 Rice gene
                    for aspartic protease (302 to 448) 147 nt"
repeat_region      complement(6435. .6517)
                    /rpt_family="AT_rich"
repeat_region      complement(7038. .7063)
                    /rpt_family="AT_rich"
repeat_region      complement(8244. .8263)
                    /rpt_family="(TGG)n"
mRNA               join(<8312. .9043,9241. .9296,9519. .>9552)
                    /gene="OSJNBa0027P10.12"
gene               8312. .9552
                    /gene="OSJNBa0027P10.12"
                    /note="similar to ethylene responsive element binding
                    protein EREBP 4 GB:BAA07323 GI:1208497 (Nicotiana
                    tabacum)"
CDS               join(8312. .9043,9241. .9296,9519. .9552)
                    /gene="OSJNBa0027P10.12"
                    /codon_start=1
                    /product="putative ethylene-responsive element binding
                    protein"
                    /protein_id="AAG60182.1"
                    /db_xref="GI:12597874"
                    /translation="MAGFGLDQHLDLIRAHLLLEDAHHHVLAPSPSPSPPGTGRVRPAP
                    VSLPPRPPLLWAAASAAPRQEECFELGGGYAGEGEGEEEDFRRYRGVRQRPWGKYA
                    AEIRDPARKGARVWLGTYDTAVEAARAYDRAAFQLRGSKAILNFPNEVAADA AVKWAP
                    PVAPIPAAAMSAGRGKRVRSEEQYYLREVKKERLIMAPPENSSSSSSSSAAAAAGDIWD
                    ELKGICSLPPLSPLSPHPMAFPQLFVIDLAFGQILNSFVLLLLLRTDFAFDYAI"
repeat_region      8539. .8564
                    /rpt_family="(AGGGGG)n"
mRNA               join(<11624. .11732,11854. .11938,13068. .13140,13236.
.13256,
                    13343. .13396,13879. .13929,14124. .14186,15136. .15201,
                    15308. .15346,16032. .16154,16674. .16766,16873. .16954,
                    17067. .17419)
                    /gene="OSJNBa0027P10.1"
gene               11624. .17419
                    /gene="OSJNBa0027P10.1"
                    /note="similar to palmitoyl protein thioesterase
                    GB:AAA85337 GI:1160967; EST AU065981 from this gene"
CDS               join(11624. .11732,11854. .11938,13068. .13140,13236.
.13256,

```

13343. .13396,13879. .13929,14124. .14186,15136. .15201,  
 15308. .15346,16032. .16154,16674. .16766,16873. .16954,  
 17067. .17095)  
 /gene="OSJNBa0027P10.1"  
 /codon\_start=1  
 /product="putative palmitoyl-protein thioesterase"  
 /protein\_id="AAG60184.1"  
 /db\_xref="GI:12597876"  
 /translation="MAYALRGAALVGVLLLVVASPALVPVASAVPFIVLHGIGDQCEN  
 GGMASFTEMLGEWSGSKGYCILNVFLLSEIGRGAWDSWLMPLQEQAQTVCKKVKMKME  
 LRKGYSIVGLSQGNLIGRAVIEYCDGGPPVKNFISIGGPHAGTASVPLCGSGIVCVLI  
 DALIKLEIYSNYVQAHLAPSGYLKIPTDMDTYLKGCKFLPKLNNEIPSERNATYKQRF  
 SSLENLVLIMFEDDAVLIPRETAWFGYYPDGAFFSPVQPPQKTKLYTEDWIGLKALEEA  
 GRVKFVSVPGEEGVPSDQF"  
 repeat\_region 13558. .13589  
 /rpt\_family="(TAAAAA)n"  
 repeat\_region complement(15485. .15668)  
 /rpt\_family="Wanderer\_Os5 MITE element from gb:X13679  
 Oryza sativa H3 histone pseudogene H3R-12 (2 to 187) 186  
 nt"  
 repeat\_region complement(15516. .15571)  
 /rpt\_family="AT\_rich"  
 repeat\_region 15563. .15647  
 /rpt\_family="Wanderer\_Ol1 MITE element from gb:U34601  
 Oryza sativa wanderer mobile element linked to Xa21 (411  
 to 637) 227 nt"  
 repeat\_region complement(17910. .17957)  
 /rpt\_family="(CGG)n"  
 mRNA join(<19796. .21016,21710. .21802,22214. .22341,  
 22644. .>22743)  
 /gene="OSJNBa0027P10.13"  
 gene 19796. .22743  
 /gene="OSJNBa0027P10.13"  
 /note="predicted by fgenesh"  
 CDS join(19796. .21016,21710. .21802,22214. .22341,22644.  
 .22743)  
 /gene="OSJNBa0027P10.13"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AAG60187.1"  
 /db\_xref="GI:12597879"  
 /translation="MSDEFDNSIDPAEIYTTDMFMAEHSVLNSFAGRIDRRIKARLEG  
 GTSRRTSGPRKYINRNHEGAHQDLFADYFAEDPLYSAATFRFRFRMRHVFLHIVDEL  
 GKWSSYFTHRVDCTGCLGHSPLQKCTAAIRMLAYGTAADTLDEYLVKVPQSTALECLEN  
 FVEGVVEVFSSRYLRPTAEDLERLLQVGESRGFPGMLGSDCMHWRWKNCPATAWKQ  
 YTRGDQKYPTIILEAVASYDLHIWHAFFGIPGSNNDINVLNQSPLEIAIKGEAPQIQ  
 FIVNGTQYNTGYLLADGIYPEWAAAFVKSIRSPQLEKHKLFAREQEGKRKDIERAFGLV  
 QARFNIVHRPARSWSQKVLKIMQACVILHNMIVEDEGEAEDPIDLNAAPGTSIVLP  
 PEVHAGSNDHPSFSMDCTIYELAVVSTFEGGLTKISAVQNDSSVVRMAGECRRDRRCRV  
 KNPMPPKKKTMLGLMAPQTVDEGTDQGDVNGSSIFLAGGEGSHSPSPKYSTKHQFSLYTL  
 QIPET"  
 repeat\_region complement(21831. .21851)  
 /rpt\_family="AT\_rich"  
 repeat\_region complement(24477. .24506)  
 /rpt\_family="AT\_rich"  
 mRNA complement(join(<25322. .25620,26566. .26629,27116. .27190,  
 27868. .27962,28341. .28428,28538. .28665,29324. .29455,



```

29557. .29637,29724. .29882,29980. .>31398))
/gene="OSJNBa0027P10.11"
gene complement(25322. .31398)
/gene="OSJNBa0027P10.11"
/note="similar to arm repeat protein GB:AAF33245
GI:6959880 (Drosophila melanogaster); EST D22130 from this
gene"
CDS complement(join(25606. .25620,26566. .26629,27116. .27190,
27868. .27962,28341. .28428,28538. .28665,29324. .29455,
29557. .29637,29724. .29882,29980. .31398))
/gene="OSJNBa0027P10.11"
/codon_start=1
/product="putative arm repeat protein"
/protein_id="AAG60190.1"
/db_xref="GI:12597882"
/translation="MTRRVRRRLCKDGGKGKDVAADEERELVSCSSSSRRRGGLGVAV
AARGGGGGSGSCVVDWRTLPPDPTVLQLFGRNLNYRDRASMAAACRTWRDLGASPLWS
ALDLRAHRCDAEVASSLSSRCGSLRRLRLRGHEAAAAASGLRARGLREVVDGCRGLT
DATLAVLAARHEALESQIGPDLERISSDALRQVAFCCSRLRRLRLSGLRDADADAI
GALARYCPLLEDVAFDCGSVDEAAIAGILSLRFLSVAGCHNLKWATASTSWAQLPSL
VAVDVSRSTDVSPSAISRLISHSKTLKLICTLNCKSVEEEEQAHNPGAFSNSKGKLVLTII

```

Query Match 7.9%; Score 33.6; DB 8; Length 141307;  
Best Local Similarity 61.4%; Pred. No. 21;  
Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```

Qy      165 gatggaaggggtggatacaactctgttcccatcaacggcggtggcagcccctagctaggc 224
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 99091 GGTGAAGGTGACGGAGCCAACACCCTTGATCTCCACGCCGGAGGCGTCCCCGAACTTGAC 99150

Qy      225 ggtggatccgagcctgtatcagaaatcg 252
      | | | | | | | | | | | | | | | |
Db 99151 GGTGCCTCGGACGCTGGAGTCGAACTCG 99178

```

```

RESULT 13
AP002836
LOCUS AP002836 146921 bp DNA PLN 26-JAN-2001
DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone:P0512G09.
ACCESSION AP002836
VERSION AP002836.1 GI:9711819
KEYWORDS .
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:P0512G09.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 146921)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0512G09
JOURNAL Published Only in DataBase (2000) In press
REFERENCE 2 (bases 1 to 146921)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-2000) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai

```

LOCUS	AP002836	146921 bp	DNA	PLN	26-JAN-2001
DEFINITION	Oryza sativa genomic DNA, chromosome 1, PAC clone:P0512G09.				
ACCESSION	AP002836				
VERSION	AP002836.1 GI:9711819				
KEYWORDS	.				
SOURCE	Oryza sativa (cultivar:Nipponbare) DNA, clone:P0512G09.				
ORGANISM	Oryza sativa				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;				
	Ehrhartoideae; Oryzeae; Oryza.				
REFERENCE	1 (bases 1 to 146921)				
AUTHORS	Sasaki,T., Matsumoto,T. and Yamamoto,K.				
TITLE	Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC				
	clone:P0512G09				
JOURNAL	Published Only in DataBase (2000) In press				
REFERENCE	2 (bases 1 to 146921)				
AUTHORS	Sasaki,T., Matsumoto,T. and Yamamoto,K.				
TITLE	Direct Submission				
JOURNAL	Submitted (03-AUG-2000) Takuji Sasaki, National Institute of				
	Agrobiological Resources, Rice Genome Research Program; Kannondai				

ACCESSION AP002836  
VERSION AP002836.1 GI:9711819

VERSION            AP002836.1    GI:9711819

KEYWORDS .  
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:P0512G09.  
ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE	1 (bases 1 to 146921)
AUTHORS	Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE	Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC clone:P0512G09

JOURNAL Published Only in DataBase (2000) In press

REFERENCE 2 (bases 1 to 146921)  
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.

TITLE Direct Submission  
JOURNAL Submitted (03-AUG-2000) Takuji Sasaki, National Institute of  
Agrobiological Resources, Rice Genome Research Program; Kannondai

2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,  
Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT

Genes were predicted from the integrated results of the following:  
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor  
(October 1998 version). The genomic sequence was searched against  
NCBI NonRedundant Protein database, nr  
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at  
RGP. Protein homologies of the coding regions were searched against  
NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent  
the identified cDNA sequences using BLASTN 2.0 with the  
corresponding DDBJ accession no. and RGP clone ID.

A gene with identity or significant homology to a protein is  
classified based on the protein name to indicate the homology level  
such as same name, 'putative-' and '-like protein'. A gene without  
significant homology to any protein but with EST homology (covering  
almost the entire length of partial sequence) is classified as an  
'unknown' protein. A gene predicted with a gene prediction program  
is classified as a 'hypothetical' protein.

The orientation of the sequence is from T7 to SP6 of the PAC clone.  
This sequence of P0512G09 clone has an overlap with P0695A04 (DDBJ:  
AP002816) clone at the 5' end and an overlap with P0710E05 (DDBJ:  
AP002743) at the 3' end. The sequence of this clone starts at the  
position 46972 of P0695A04 and ends at the position 31325 of  
P0710E05. Detailed information on overlap and assembly quality  
together with annotation of this entry is available at  
<http://rgp.dna.affrc.go.jp/GenomeSeq.html>.

FEATURES

source

Location/Qualifiers

1. .146921  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/chromosome="1"  
/clone="P0512G09"

gene

join(5036. .5129,5308. .5485,5682. .5763,5786. .7147)  
/gene="P0512G09.1"

CDS

join(5036. .5129,5308. .5485,5682. .5763,5786. .7147)  
/gene="P0512G09.1"

/note="contains ESTs  
AU070282 (S21326), C26605 (C12676), C72475 (E1678),  
D15872 (C1448), AU091541 (C12676)  
unknown protein"  
/codon\_start=1  
/protein\_id="BAB07919.1"  
/db\_xref="GI:9711820"  
/translation="MPRYVAADCGRSNVAQTLVNPCAITLPSKIKAPPVTEESPYPLS  
LPCGPMTRGHHQTTTIPTNQTTETRRRRRQRQRRGSILLASRSIRDVTPSDQPSSIIAK  
PTSEISSIVVVVAVEMSSLTAAQFRSLFEMAGGEDISFLFGDDEAAVAAPVELQ GK  
RGWEEVDQGE GSGAVA AKRQRSPTSSRENSSGSNEGGQEEVSEAAAAMAAVGRGGGR  
RLWVKERDSEWDMVSSPDYPDSEFRKAFRMSKATFEVVCDELAAVAKEDTMLRAAI  
PVRKRVAVCVWRLATGEPLRLVSKRFLGISTCHKLVLEVCAALKAMVMPKVVRWPEA  
GDAAAIAAHFEAISGISGVVGAITYTHIPIIAPKSNVASYNNRRHTERNQKTSYSMTV  
QCVVDSTGAFTDVCIGWPGSNSDEEVLEKSALYLHRGVPGLIQGWVVGGSFPLMDW  
MLVPYTHQNLTWAQHMLNEKVA AVRGVARDAFERLKRRWGCLQKRTEVKLLDLPTVLG  
ACCVLHNICERSGDAVVDADDCAFDLFDDDMVAENAVRSSAAAQARDAIAHNLLHSGG  
GASFF"

gene

join(9440. .9958,10386. .10709)

CDS /gene="P0512G09.2"  
 join(9440. .9958,10386. .10709)  
 /gene="P0512G09.2"  
 /note="contains ESTs AU067976(C11383),AU067977(C11383)  
 unknown protein"  
 /codon\_start=1  
 /protein\_id="BAB07920.1"  
 /db\_xref="GI:9711821"  
 /translation="MADDGDGDCVNLEPFFYDEAATVAEAAAAAERREREEQEKAREA  
 AANARRWAAHNAALAGIREYDPAEETIYTRYHYADLSEFDLDEESRLPPMRHTAATY  
 APPARALHFLCDMINVLAVRIILPSSDRSDGGGVGFPISVYGSVIARDQLDYKCVHLF  
 RRCRDDPQLITSEDELSLILTGPHRGLVLYDALYIEVDLKMVKVKGDDQQGCKDKRLSK  
 GLIVLDGVLLSTNLSDDLRAAVKTATLDRRSTMPCAVQVTYAYVTRAVEATVSVELLH  
 DQGG"  
 gene complement(join(14232. .14299,15647. .15798,18978. .18994))  
 /gene="P0512G09.3"  
 CDS complement(join(14232. .14299,15647. .15798,18978. .18994))  
 /gene="P0512G09.3"  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="BAB07921.1"  
 /db\_xref="GI:9711822"  
 /translation="MSGIYSACGVASVAAGSGRGRATVPLTSFNAWRRRVGHYRSPQTA  
 PEGVEGGRERREVRIRRGEGESESDSKADTRGTV"  
 gene join(19202. .19243,19551. .19874)  
 /gene="P0512G09.4"  
 CDS join(19202. .19243,19551. .19874)  
 /gene="P0512G09.4"  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="BAB07922.1"  
 /db\_xref="GI:9711823"  
 /translation="MGRRWRLGFSPPAQTKHVRVFSSGTDHQLLRSLVNPGEEVILTV  
 PNEQLEHVVRFLLVTRITRIMTSDDVLVRS LGNAYFLVPTMVNLHPTLAAARLDGRVK  
 VSSACGAAKACSVGRLSRK"  
 gene complement(join(20820. .20996,21294. .21316,22118. .22199,  
 22461. .23639))  
 /gene="P0512G09.5"  
 CDS complement(join(20820. .20996,21294. .21316,22118. .22199,  
 22461. .23639))  
 /gene="P0512G09.5"  
 /note="hypothetical protein  
 similar to Oryza sativa chromosome 6, P0675A05.27"  
 /codon\_start=1  
 /protein\_id="BAB07923.1"  
 /db\_xref="GI:9711824"  
 /translation="MKAILCVFEHLSGLKINFHKSEIFCFGAKEMEDQYRQLFGCNS  
 GSFPFRYLGIPHYRKLNSDWKCVEDMFQRKLSTWKGKNNSYGGRLVLLNSVLSSLP  
 MFMLLFFEIPRGVLQRLDYRSSFWSQKKKYRLTKWDYIYRLKDQGGGLGVNLND  
 IMNRCLLSKWLFKLLNGEGLWQNLRLNRYLKGKPLSHMSHKPGTSQFWAGLMKVVDQF  
 FQYGSFKPGNGMEIRFWEDTWLGLQPLKYQPSLYNVVRKRHSTLVEVMSTTPLNVSF  
 RRSIVGPKLVEWNDLISRLANITLSNEKDCFIWSLYKNGHFSVKS MYNAIINSNVITH  
 KRILWKVKVPLKIKVFMWFLHKKVILTKDNLIKRKWRGNKQCCFCNTQETIQHFFDC  
 HEQQDTMSNGATRLKLVAKELLQFGWRVAVSEQLAKEQPSPLPYIVDYTGETLLES  
 R KDELLKGAMEINYLGDVGVVVCRRVFDVFLSHWKV"  
 gene join(25126. .25245,25311. .25409,25606. .25740)  
 /gene="P0512G09.6"

```

CDS
join(25126. .25245,25311. .25409,25606. .25740)
/gene="P0512G09.6"
/note="contains ESTs C26486(C12424),AU091537(C12424)
unknown protein"
/codon_start=1
/protein_id="BAB07924.1"
/db_xref="GI:9711825"
/translation="MISSSEEATTSYLASCLTLLGLCFCVLVGVVRRKRGRVRELCLV
TSLSVSHLISLSLFLPHVLYCSLISIFFEISMFGGLTEESEHTRCFLSMQEPWCWELC
VLTLYLSAVHHAQYD"

gene
.36317,
join(32826. .33658,34519. .34634,34939. .35147,36198.

36430. .36558,36639. .36764,37699. .37754,37853. .37952,
38043. .38160,38577. .38786,38875. .39187,39784. .39877,
40099. .40207,40488. .40759,40842. .40969,41077. .41311,
41815. .42003,42090. .42373,42532. .42643,42731. .42973,
43149. .43231,43364. .43511)
/gene="P0512G09.7"

CDS
.36317,
join(32826. .33658,34519. .34634,34939. .35147,36198.

36430. .36558,36639. .36764,37699. .37754,37853. .37952,
38043. .38160,38577. .38786,38875. .39187,39784. .39877,
40099. .40207,40488. .40759,40842. .40969,41077. .41311,
41815. .42003,42090. .42373,42532. .42643,42731. .42973,
43149. .43231,43364. .43511)
/gene="P0512G09.7"
/note="contains EST C72119(E1022)
unknown protein"
/codon_start=1
/protein_id="BAB07925.1"
/db_xref="GI:9711826"
/translation="MAPHLHLRLLLLVVGVALGSAGLGWGGGGGEGEATDREAPDPYSI
LTWHDYSPSPPPPPPPVAPAATCEGDLHGKGNFSTRCEVSEEEVLGGDVYITGEGS
LVLLAGAALTCQRPQCVISANLSGEVRLGRGVRIAGRVSLAAANVTIADTVVNTTA
LAGDPPERTSGVPTGTHGDGGGHGGRGASCYVKDQGTQEDSWGGDAYAWSDLHPFSY
GSKGGSTSVKDYGGSGGGIVWLYADDLIMNGTVLADGDSSEKGGGGSGGSIYIKSK
TMHGAGKISASGGNLAGGGGGRVSINVFSRHDDTQVFAHGGKSSGCPDNAGAAGTLY
EAVPKSLVVSNNNLSTQTDTLLEFPNQPLWTNVFVKNHAKVAVPLLWSRVQVQGQLS
LLSGAILTFGLTRYPYSEFELMAEELLMSDSTIKVFGALRMSVKMLLMWNSKMLIDGG
GDSIVATSLLDASNLIVLKESSVIHSNANLGVRGQGLLNLSGEGDIEAQRILISLFY
SIKVGPGSILRGPLVNGSSGDVAPKLNCDDDICPVEIIHPPEDCNLNSSLSFTLQVCR
VEDIDIWGLVQGTVIHFNRARSVSVHTSGTISATGLGCRSGVGQKILNSGVSGGGGH
GGRGGDGFYNESHAEGGSMYGSADLPCELGSGSGNDTTKLSTAGGGIIVMGSWEYSLP
SLSLYGSVESNGQSSTDVVTNASIGGPGGGSGGTILLFVRALS LAESSILSSVGGLGN
FGSGGGGGGRIHFHWSNIPTGDEYVPVAAVKGSIRTSGGISKGKGFPGENGTVTGKAC
PKGLYGTFCKECPLGTYKNVTGSSKSLCVQCPPDELPHRAIYTSVRGGAYETPCPYKC
VSDRYRMPHCYTALEELIYTFGGPWLFGLLLSGLLVLLALVLSVARMKFVGTDELPGP
APTQQGSQIDHSFPFLESLENEVLETNRAEESHGHVHRMYFMGPNTFSEPWHLPHTPPE
QISEIVYEDAFNRVDEINTLAAYQWWEHSIHSILCVLAYPLAWSWQQFRRRKKLQRL
REFVRSEYDHSCLRSCRSRALYEGLKVTATPDLMLGYLDFFLGDEKRPDLPPRLRQR
FPMCLIFGGDGSYMAPFSLHSDSVLTSLMSQAVPSSIWHRLVAGLNAQLRLVRRGSLR
GTFLPVLDWLETHANPSLGVNGVRVDLAWFQATALGYCQLGLVYVAVEEPMSAELDGS
PRIKIEQHSLITGGILDSNSLRTLKDRDLFPFSLILHNTKPVGHQDLVGLVYSILL
LADFSLVLLTFLQLYSYSMADVLLVLFVLPLGILSPFPAGINALFSHGPRRSAGLARV
YALWNITSLNVVVAFACGLVHYKSSTKRHPSTQPWNLTDES GWLFP TGLMLLKCI
QARLVDWHVANLEIQDRAVYSNDPSIFWQS"

```

gene join(44868. .44915,45001. .45052,47765. .47937,48241.  
.48336,  
48920. .49009,49126. .49374,49478. .49657,51850. .51950,  
52039. .52204)  
/gene="P0512G09.8"  
CDS join(44868. .44915,45001. .45052,47765. .47937,48241.  
.48336,  
48920. .49009,49126. .49374,49478. .49657,51850. .51950,  
52039. .52204)  
/gene="P0512G09.8"  
/note="hypothetical protein  
similar to Arabidopsis thaliana, M7J2.190"  
/codon\_start=1  
/protein\_id="BAB07926.1"

Query Match 7.9%; Score 33.6; DB 8; Length 146921;  
Best Local Similarity 61.4%; Pred. No. 21;  
Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 165 gatggaaggggtggatacaactctgttcccatcaacggcggtggcagccctagctaggc 224  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 138989 GGTGAAGGTGACGGAGCCAACACCCTTGATCTCCACGCCGAGGCGTCCCCGAAGTAC 139048  
Qy 225 ggtggatccgagcctgtatcagaaatcg 252  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 139049 GGTGCCTCGGACGCTGGAGTCGAAGTTCG 139076

RESULT 14  
AC024105/c  
LOCUS AC024105 162700 bp DNA HTG 06-SEP-2000  
DEFINITION Homo sapiens chromosome 12 clone RP11-571I17, WORKING DRAFT  
SEQUENCE, 14 unordered pieces.  
ACCESSION AC024105  
VERSION AC024105.13 GI:9966537  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 162700)  
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,  
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,  
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,  
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,  
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,  
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,  
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hagues,M.,  
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,  
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,  
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,  
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,  
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,  
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,  
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,  
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,  
Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,

Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,  
Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and  
Gibbs,R.

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 162700)  
AUTHORS Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (24-FEB-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
COMMENT On Sep 5, 2000 this sequence version replaced gi:9438319.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: HAHT  
Center clone name: RP11-571I17

----- Summary Statistics  
Sequencing vector: M13; L08821  
Chemistry: Dye-terminator Big Dye: 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 144080 bases at least Q40  
Consensus quality: 155185 bases at least Q30  
Consensus quality: 158367 bases at least Q20  
Estimated insert size: 159112; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 35277: contig of 35277 bp in length  
\* 35278 35377: gap of unknown length  
\* 35378 50496: contig of 15119 bp in length  
\* 50497 50596: gap of unknown length  
\* 50597 69110: contig of 18514 bp in length  
\* 69111 69210: gap of unknown length  
\* 69211 82030: contig of 12820 bp in length  
\* 82031 82130: gap of unknown length  
\* 82131 94011: contig of 11881 bp in length  
\* 94012 94111: gap of unknown length  
\* 94112 104185: contig of 10074 bp in length  
\* 104186 104285: gap of unknown length  
\* 104286 115314: contig of 11029 bp in length  
\* 115315 115414: gap of unknown length  
\* 115415 124235: contig of 8821 bp in length  
\* 124236 124335: gap of unknown length

\* 124336 133848: contig of 9513 bp in length  
 \* 133849 133948: gap of unknown length  
 \* 133949 143516: contig of 9568 bp in length  
 \* 143517 143616: gap of unknown length  
 \* 143617 151726: contig of 8110 bp in length  
 \* 151727 151826: gap of unknown length  
 \* 151827 157897: contig of 6071 bp in length  
 \* 157898 157997: gap of unknown length  
 \* 157998 161083: contig of 3086 bp in length  
 \* 161084 161183: gap of unknown length  
 \* 161184 162700: contig of 1517 bp in length.

FEATURES Location/Qualifiers  
 source 1. .162700  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="12"  
 /clone="RP11-571I17"

BASE COUNT 46705 a 33082 c 32731 g 48849 t 1333 others  
 ORIGIN

Query Match 7.9%; Score 33.6; DB 2; Length 162700;  
 Best Local Similarity 52.1%; Pred. No. 21;  
 Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 53 ctcccctgtgctctacttctgcctgatggcccttgctcgtagctgctatggctctgtgtcat 112  
 ||||| | | | | |||| | ||| | || ||||| ||||| |||||  
 Db 43901 CTCCTCGGGTCACCGCTTCACTCCCCTGGGACCTGATGTAACATATCATGATCAATGTCAT 43842  
 Qy 113 gtacaccacctcggcacaagcaggaaggagtggctacaactcgtacgaacctgatggaag 172  
 ||||| | | | | ||| | |||| | ||| || ||||  
 Db 43841 GTACAGCCTTCCAGATCTTCTCCTGAGGTTTTGCTAACATATTTACCCACGGATAGGAAT 43782  
 Qy 173 gggtggatacaactctgttcccat 196  
 || | ||| || | || | |  
 Db 43781 TGGAGCATATAATGTTTTTGTCTT 43758

RESULT 15  
 AC074345/c  
 LOCUS AC074345 178141 bp DNA HTG 29-AUG-2000  
 DEFINITION Homo sapiens chromosome 12 clone RP11-349K16, WORKING DRAFT  
 SEQUENCE, 23 unordered pieces.  
 ACCESSION AC074345  
 VERSION AC074345.3 GI:9937816  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 178141)  
 AUTHORS Waterston,R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 178141)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission

JOURNAL Submitted (28-JUL-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

COMMENT On Aug 29, 2000 this sequence version replaced gi:9665208.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----

Center project name: H\_NH0349K16

----- Summary Statistics -----

Sequencing vector: M13; 100%

Sequencing vector: plasmid; 0%

Chemistry: Dye-primer ET; 100% of reads

Chemistry: Dye-terminator Big Dye; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 165955 bases at least Q40

Consensus quality: 169539 bases at least Q30

Consensus quality: 171386 bases at least Q20

Insert size: 173000; agarose-fp

Insert size: 175941; sum-of-contigs

Quality coverage: 4.39 in Q20 bases; agarose-fp

Quality coverage: 4.46 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 23 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\*       1       3373: contig of 3373 bp in length  
\*       3374       3473: gap of unknown length  
\*       3474       8739: contig of 5266 bp in length  
\*       8740       8839: gap of unknown length  
\*       8840       14233: contig of 5394 bp in length  
\*       14234       14333: gap of unknown length  
\*       14334       20608: contig of 6275 bp in length  
\*       20609       20708: gap of unknown length  
\*       20709       25795: contig of 5087 bp in length  
\*       25796       25895: gap of unknown length  
\*       25896       30735: contig of 4840 bp in length  
\*       30736       30835: gap of unknown length  
\*       30836       39879: contig of 9044 bp in length  
\*       39880       39979: gap of unknown length  
\*       39980       47786: contig of 7807 bp in length  
\*       47787       47886: gap of unknown length  
\*       47887       55623: contig of 7737 bp in length  
\*       55624       55723: gap of unknown length  
\*       55724       65472: contig of 9749 bp in length  
\*       65473       65572: gap of unknown length  
\*       65573       75441: contig of 9869 bp in length  
\*       75442       75541: gap of unknown length  
\*       75542       85314: contig of 9773 bp in length  
\*       85315       85414: gap of unknown length



```

*      85415      95969: contig of 10555 bp in length
*      95970      96069: gap of unknown length
*      96070     109739: contig of 13670 bp in length
*     109740     109839: gap of unknown length
*     109840     123818: contig of 13979 bp in length
*     123819     123918: gap of unknown length
*     123919     141676: contig of 17758 bp in length
*     141677     141776: gap of unknown length
*     141777     159319: contig of 17543 bp in length
*     159320     159419: gap of unknown length
*     159420     160861: contig of 1442 bp in length
*     160862     160961: gap of unknown length
*     160962     163347: contig of 2386 bp in length
*     163348     163447: gap of unknown length
*     163448     165891: contig of 2444 bp in length
*     165892     165991: gap of unknown length
*     165992     169994: contig of 4003 bp in length
*     169995     170094: gap of unknown length
*     170095     174708: contig of 4614 bp in length
*     174709     174808: gap of unknown length
*     174809     178141: contig of 3333 bp in length.

```

```

FEATURES      Location/Qualifiers
source      1. .178141
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="12"
             /clone="RP11-349K16"
misc_feature 1. .3373
             /note="assembly_name:Contig10
             clone_end:T7
             vector_side:left"
misc_feature 3474. .8739
             /note="assembly_name:Contig11"
misc_feature 8840. .14233
             /note="assembly_name:Contig12"
misc_feature 14334. .20608
             /note="assembly_name:Contig13"
misc_feature 20709. .25795
             /note="assembly_name:Contig14"
misc_feature 25896. .30735
             /note="assembly_name:Contig15"
misc_feature 30836. .39879
             /note="assembly_name:Contig16"
misc_feature 39980. .47786
             /note="assembly_name:Contig17"
misc_feature 47887. .55623
             /note="assembly_name:Contig18"
misc_feature 55724. .65472
             /note="assembly_name:Contig19"
misc_feature 65573. .75441
             /note="assembly_name:Contig20"
misc_feature 75542. .85314
             /note="assembly_name:Contig21"
misc_feature 85415. .95969
             /note="assembly_name:Contig22"
misc_feature 96070. .109739
             /note="assembly_name:Contig23"

```

BASE COUNT	51733	a	35937	c	36052	g	52218	t	2201	others
ORIGIN										

Qy	53	ctccccgtgctctacttctgcctgatggcccttgctcgtagctgctatggtctgtgtcat	112
Db	11515	CTCCTCGGGTCACCGCTTCACTCCCCTGGGACCTGATGTAACATATCATGATCAATGTCAT	11456
Qy	113	gtacaccacctcggcacaagcaggaaggagtggctacaactcgtacgaacctgatggaag	172
Db	11455	GTACAGCCTTCCAGATCTTCTCCTGAGGTTTTGCTAACATATTTACCCACGGATAGGAAT	11396
Qy	173	gggtggatacaactctgttcccat	196
Db	11395	TGGAGCATATAATGTTTTTGTCTT	11372

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

```
Title:          US-09-394-745-6332
Perfect score:  425
Sequence:       1 cggacgcgctgggtgcaattt.....tgtggtgcctctctcaacct 425
```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101:\*

- 1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:\*
- 2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:\*
- 3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:\*
- 4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:\*
- 5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:\*
- 6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:\*
- 7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:\*
- 8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:\*
- 9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:\*
- 10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:\*
- 11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:\*
- 12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:\*
- 13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:\*
- 14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:\*
- 15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:\*
- 16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:\*
- 17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:\*
- 18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:\*
- 19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:\*
- 20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:\*
- 21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:\*
- 22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query		Length	DB	ID	Description
		Match					
c	1	32.6	7.7	1559	21	AAC69485	Human secreted pro
	2	32.4	7.6	565	22	AAH99273	Human protein enco
	3	32.4	7.6	4360	22	AAD09494	Human SGP018 phosp
	4	31.4	7.4	545	21	AAF12056	Aspergillus oryzae
	5	30.4	7.2	2075	13	AAQ20760	Rice light-harvest
	6	30.2	7.1	3326	22	AAF74625	Human GLI-1 nucleo
	7	30.2	7.1	3600	22	AAD12302	Human Cubitus inte
	8	30	7.1	5118	21	AAC44700	Arabidopsis thalia
	9	29.8	7.0	1229	12	AAQ14640	Plasmid pGB18ARR i
	10	29.8	7.0	1230	17	AAT34381	Plasmid pGB18ARR i

	11	29.8	7.0	1230	20	AAZ32251	Human temporal lob
	12	29.8	7.0	1230	21	AAA88186	pGB18ARR human tem
	13	29.8	7.0	1481	20	AAZ32285	Nucleotide sequenc
	14	29.8	7.0	2702	17	AAT34373	Plasmid pJC44x (AT
	15	29.8	7.0	2702	20	AAZ32236	Human glioblastoma
	16	29.8	7.0	2702	21	AAA88171	pJC44x human gliob
	17	29.8	7.0	3131	12	AAQ14630	Plasmid pTM3 inser
	18	29.8	7.0	3131	17	AAT34374	Plasmid pTM3 (ATCC
	19	29.8	7.0	3131	21	AAA88176	pTM3 human gliobla
	20	29.8	7.0	3705	20	AAZ32265	Human dunce-like p
	21	29.8	7.0	3705	21	AAA88200	pPDE46 human dunce
	22	29.6	7.0	740	19	AAV61151	5' cDNA sequence o
	23	29.6	7.0	740	19	AAV58494	5' fragment of pro
	24	29.6	7.0	740	21	AAA06257	Human immunogenic
	25	29.6	7.0	740	22	AAS10016	Human prostate tum
	26	29.6	7.0	740	22	AAH93373	Human prostate-spe
	27	29.6	7.0	740	22	AAH84687	Human prostate-spe
	28	29.6	7.0	740	22	AAH02438	Prostate tumour an
c	29	29.6	7.0	1642	22	AAH99665	Human protein enco
c	30	29.6	7.0	2442	20	AAX58259	Vascular endotheli
c	31	29.6	7.0	2442	21	AAA89049	Human tumour necro
c	32	29.6	7.0	2442	21	AAZ50754	Human TNF-gamma-al
c	33	29.6	7.0	2683	20	AAX58249	Vascular endotheli
	34	29.4	6.9	5503	21	AAC74570	Human ORFX ORF125
c	35	29.2	6.9	2842	21	AAD00334	Rice raffinose syn
	36	29	6.8	423	21	AAH30837	Human colon cancer
c	37	28.8	6.8	569	22	AAI42639	Probe #11325 used
c	38	28.8	6.8	1077	21	AAZ94582	Maize cyclin D ZmC
	39	28.8	6.8	1083	22	AAS08158	Mammalian toxicolo
c	40	28.8	6.8	1846	19	AAV33889	Maize CYCD2 gene.
c	41	28.8	6.8	1861	21	AAZ94581	Maize cyclin D ZmC
	42	28.6	6.7	498	20	AAZ22437	Internal Transcrib
	43	28.6	6.7	556	20	AAZ22438	Internal Transcrib
	44	28.4	6.7	251	19	AAX11226	Human biallelic po
	45	28.4	6.7	427	21	AAZ42280	Human 5' EST isola

#### ALIGNMENTS

RESULT 1

AAC69485

ID AAC69485 standard; cDNA; 1559 BP.

XX

AC AAC69485;

XX

DT 30-JAN-2001 (first entry)

XX

DE Human secreted protein gene 31 SEQ ID NO:41.

XX

KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;

KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;

KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;

KW fungicide; ophthalmological; gene therapy; autoimmune disease; infection;

KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;

KW cerebrovascular disorder; nervous system disorder; ocular disorder;

KW wound healing; skin aging; food additive; preservative; ss.

XX  
 OS Homo sapiens.  
 XX  
 PN WO200058469-A1.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 23-MAR-2000; 2000WO-US07579.  
 XX  
 PR 26-MAR-1999; 99US-0126509.  
 PR 07-JAN-2000; 2000US-0174853.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM, Komatsoulis G;  
 XX  
 DR WPI; 2000-594642/56.  
 DR P-PSDB; AAB38233.  
 XX  
 PT Isolated nucleic acid molecule encoding a human secreted protein is  
 PT used in preventing, treating or ameliorating a medical condition -  
 XX  
 PS Claim 1; Page 348; 416pp; English.  
 XX  
 CC The polynucleotide sequences given in AAC69455 to AAC69502 encode the  
 CC human secreted proteins given in AAB38203 to AAB38250. AAB38251 to  
 CC AAB38320 represent human secreted polypeptide sequences and proteins  
 CC homologous to them, which are given in the exemplification of the present  
 CC invention. Human secreted proteins have activities based on the tissues  
 CC and cells the genes are expressed in. Example of activities include:  
 CC immunosuppressive; antiarthritic; antirheumatic; antiproliferative;  
 CC cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;  
 CC neuroprotective; antibacterial; virucide; fungicide; and  
 CC ophthalmological. The polynucleotides and polypeptides can be used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
 CC in diagnosing a pathological condition or susceptibility to a  
 CC pathological condition. Disorders which are diagnosed or treated include  
 CC autoimmune diseases, hyperproliferative disorders, cardiovascular  
 CC disorders, cerebrovascular disorders, angiogenesis, nervous system  
 CC disorders, infections caused by bacteria, viruses and fungi and ocular  
 CC disorders. The polypeptides can also be used to aid wound healing and  
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
 CC maintain organs before transplantation, for supporting cell culture of  
 CC primary tissues, to regenerate tissues and in chemotaxis. The  
 CC polypeptides can also be used as a food additive or preservative to  
 CC increase or decrease storage capabilities. AAC69446 to AAC69454 and  
 CC AAB38202 represent sequences used in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 1559 BP; 236 A; 525 C; 389 G; 405 T; 4 other;

Query Match 7.7%; Score 32.6; DB 21; Length 1559;  
 Best Local Similarity 51.7%; Pred. No. 2;  
 Matches 74; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 283 cactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaacgg 342  
 |||| | ||| | | | | | ||| ||||| || | |  
 Db 1340 cactccagctgctgcttcaggaccagatgtcgtggctgctcacgctctcccaggcgctg 1399  
 Qy 343 ctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccctcg 402  
 ||| |||| | | ||| || ||||| || ||||| ||||  
 Db 1400 ctctcgctcagggtgcgccgccgctccccaccgaggagccagcgctcgtctctctctcc 1459  
 Qy 403 ttctgtggtgcctctctcaacct 425  
 |||| || || | ||||  
 Db 1460 ttctctctctcccttccccacct 1482

RESULT 2

AAH99273

ID AAH99273 standard; cDNA; 565 BP.

XX

AC AAH99273;

XX

DT 16-OCT-2001 (first entry)

XX

DE Human protein encoding cDNA sequence SEQ ID NO:108.

XX

KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
 KW antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;  
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KW neurological disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200153455-A2.

XX

PD 26-JUL-2001.

XX

PF 22-DEC-2000; 2000WO-US35017.

XX

PR 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

DR WPI; 2001-457603/49.

DR P-PSDB; AAM25332.

XX  
PT Isolated human polynucleotides encoding polypeptides, useful for the  
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
XX  
PS Claim 1; Page 347; 1217pp; English.  
XX  
CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
CC AAM25963. The proteins can have activities based on the tissues and  
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;  
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;  
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;  
CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;  
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
CC encoding them can be used in gene therapy, antisense therapy and vaccine  
CC production, The proteins and polynucleotides are useful for screening for  
CC agonists or antagonists of a protein and for the treatment and diagnosis  
CC of disorders associated with the activity of a protein e.g. inflammation,  
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
CC neurological disorders.  
XX  
SQ Sequence 565 BP; 69 A; 221 C; 126 G; 149 T; 0 other;

Query Match 7.6%; Score 32.4; DB 22; Length 565;  
Best Local Similarity 56.6%; Pred. No. 1.5;  
Matches 60; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 320 tgcacacgctagccgattaacggctcacgctaccaggcgctctacgcggatgtgccccct 379  
||| ||||| || | | |||| |||| | | ||| ||  
Db 379 tgctcacgctctcccaggcgctgctctcgctcaggggtgcgccgccgctccccaccgagg 438  
Qy 380 agccagcttctctctccccctcggttctgtggtgcctctctcaacct 425  
|||||| || ||||| |||| |||| || || | ||||  
Db 439 agccagcgctgctctctctctctctctctctctctctctcccttccccacct 484

RESULT 3  
AAD09494/c  
ID AAD09494 standard; DNA; 4360 BP.  
XX  
AC AAD09494;  
XX  
DT 10-SEP-2001 (first entry)  
XX  
DE Human SGP018 phosphatase polypeptide encoding DNA.  
XX  
KW Human; SGP018 phosphatase polypeptide; phosphatase-related disease;  
KW immune-related disorder; ocular disease; organ transplant rejection;  
KW infection; diabetes; pain; sexual dysfunction; Alzheimer's disease;

KW metabolic disorder; haematopoietic cancer; mood disorder; cardiant;  
 KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
 KW cardiovascular disease; brain; neuronal-associated disease; dyskinesia;  
 KW attention disorder; cognition disorder; psychotic disorder; cytostatic;  
 KW neurological disorder; virucide; nootropic; cerebroprotective; therapy;  
 KW neuroprotective; antibacterial; vulnerary; tranquilliser; antiasthmatic;  
 KW hypotensive; immunosuppressive; antipsoriatic; analgesic; hypertensive;  
 KW antifungal; dual specificity phosphatase; DSP; MAP kinase phosphatase;  
 KW MKP; migraine; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 208..3609  
 FT /\*tag= a  
 FT /product= "Human SGP018 phosphatase polypeptide"  
 XX  
 PN WO200146394-A2.  
 XX  
 PD 28-JUN-2001.  
 XX  
 PF 21-DEC-2000; 2000WO-US34736.  
 XX  
 PR 21-DEC-1999; 99US-0173255.  
 PR 28-DEC-1999; 99US-0175766.  
 PR 25-JAN-2000; 2000US-0178078.  
 PR 31-JAN-2000; 2000US-0179301.  
 XX  
 PA (SUGE-) SUGEN INC.  
 XX  
 PI Plowman GD, Martinez R, Whyte D, Manning G, Sudarsanam S, Hill RJ;  
 PI Flanagan P;  
 XX  
 DR WPI; 2001-418058/44.  
 DR P-PSDB; AAE04836.  
 XX  
 PT Novel phosphatase polypeptide useful for treating cancers,  
 PT immune-related diseases and disorders, cardiovascular disease, brain or  
 PT neuronal-associated diseases and metabolic disorders -  
 XX  
 PS Claim 29; Fig 1; 186pp; English.  
 XX  
 CC The present invention relates to phosphatase polypeptides, nucleotide  
 CC sequences encoding them, as well as various products and methods useful  
 CC for the diagnosis and treatment of various phosphatase-related diseases  
 CC and conditions. Substance that modulates the activity of phosphatase  
 CC polypeptide is used to treat immune-related diseases and disorders,  
 CC cardiovascular disease, brain or neuronal-associated diseases and  
 CC metabolic disorders, including cancers of tissues, cancers of  
 CC haematopoietic origin, diseases of central and peripheral nervous  
 CC system, Alzheimer's disease, Parkinson's disease, multiple sclerosis,  
 CC amyotrophic lateral sclerosis, viral infections, infections caused by  
 CC prions, bacteria and fungi, ocular diseases, diabetes, migraines, pain,  
 CC sexual dysfunction, mood disorders, attention disorders, cognition  
 CC disorders, hypotension, hypertension, psychotic disorders, neurological  
 CC disorders, dyskinesias and organ transplant rejection. The present  
 CC sequence is a DNA encoding human SGP018 phosphatase polypeptide. This



CC sequence is classified as dual specificity phosphatase (DSP) and MAP  
CC kinase phosphatase (MKP).  
XX  
SQ Sequence 4360 BP; 1138 A; 1076 C; 1363 G; 783 T; 0 other;

Query Match 7.6%; Score 32.4; DB 22; Length 4360;  
Best Local Similarity 56.6%; Pred. No. 3.5;  
Matches 60; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 320 tgcacacgctagccgattaacggctcacgctaccaggcgctctacgcggatgtgccccct 379  
||| ||||| || | |||| |||| | ||| ||  
Db 1699 TGCTCACGCTCTCCCAGGCGCTGCTCTCGCTCAGGGTGCGCCGCCGCTCCCCACCGAGG 1640  
  
Qy 380 agccagcttctctctccccctcggttctgtggtgcctctctcaacct 425  
||||||| || ||||| |||| || || |||||  
Db 1639 AGCCAGCGTCGCTCTCCTCCTCCTTCTCCTCCTCCTTCCCCACCT 1594

RESULT 4

AAF12056

ID AAF12056 standard; cDNA; 545 BP.

XX

AC AAF12056;

XX

DT 13-MAR-2001 (first entry)

XX

DE Aspergillus oryzae EST SEQ ID NO:4579.

XX

KW Multiple gene expression; filamentous fungal cell; EST;

KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;

KW culture condition; environmental stress; spore morphogenesis;

KW metabolic pathway engineering; catabolic pathway engineering; ss.

XX

OS Aspergillus oryzae.

XX

PN WO200056762-A2.

XX

PD 28-SEP-2000.

XX

PF 22-MAR-2000; 2000WO-US07781.

XX

PR 22-MAR-1999; 99US-0273623.

XX

PA (NOVO ) NOVO NORDISK BIOTECH INC.

PA (NOVO ) NOVO NORDISK AS.

XX

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

XX

DR WPI; 2000-594572/56.

XX

PT Monitoring differential expression of genes in filamentous fungal cells

PT uses fluorescence-labeled nucleic acids isolated from the cells and a

PT substrate of expressed sequence tags -

XX

PS Claim 88; Page 1949; 3161pp; English.

XX  
 CC The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring  
 CC the global expression of genes from FF cells allows the production  
 CC potential of the microorganisms to be improved. New genes may be  
 CC discovered, possible functions of unknown open reading frames can be  
 CC identified and gene copy number variation and stability can be  
 CC monitored. The expression of genes can be used to study how FF cells  
 CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
 CC *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from *Aspergillus*  
 CC *niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and  
 CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are  
 CC all specifically claimed in the present invention.

XX  
 SQ Sequence 545 BP; 137 A; 148 C; 131 G; 129 T; 0 other;

Query Match 7.4%; Score 31.4; DB 21; Length 545;  
 Best Local Similarity 57.7%; Pred. No. 3.1;  
 Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 30 gacgagatcatgaggaagcaataactcccctgtgctctacttctgcctgatggcccttgtc 89  
 || | || | ||||| ||||| || | | | || | || ||  
 Db 155 gagctgctcaagcagaagcagtagtactccctatgtccggttccgacatgggtccctctcatc 214  
 Qy 90 gtagctgctatgggtctgtgtcatgtacaccacctcgg 126  
 | |||| | | || | | || | || | || |  
 Db 215 ttcgctggtgtcaacgggtcacctcgacaacatccccg 251

RESULT 5  
 AAQ20760  
 ID AAQ20760 standard; DNA; 2075 BP.  
 XX  
 AC AAQ20760;  
 XX  
 DT 16-APR-1992 (first entry)  
 XX  
 DE Rice light-harvesting chlorophyll a/b-combined protein gene.  
 XX  
 KW promoter; LHCP II; photosynthesis; ss.  
 XX  
 OS *Oryza sativa*.  
 XX  
 FH Key Location/Qualifiers

FT CDS 864..1661  
 FT /\*tag= a  
 FT /product= LHCP\_II  
 FT promoter 20..863  
 FT /\*tag= b  
 FT polyA\_signal 1825..1830  
 FT /\*tag= c  
 XX  
 PN JP03277291-A.  
 XX  
 PD 09-DEC-1991.  
 XX  
 PF 27-MAR-1990; 90JP-0075774.  
 XX  
 PR 27-MAR-1990; 90JP-0075774.  
 XX  
 PA (MITK ) MITSUI TOATSU CHEM INC.  
 PA (NORQ ) NORINSHO KK.  
 XX  
 DR WPI; 1992-029693/04.  
 XX  
 PT Photosynthesis-related gene, for new plant species - comprises  
 PT DNA acid fragment contg. promoter of light-collecting  
 PT chlorophyll-combined protein gene obtd. from rice plant  
 XX  
 PS Claim 3; Fig 1; 6pp; Japanese.  
 XX  
 CC LHCP II sequences were isolated from a cDNA library prepared from  
 CC total RNA of 2-week old rice shoots. The library was probed by a  
 CC 17mer probe based on part of the (known) LHCP II coding sequence.  
 CC A positive cDNA clone was then used to screen a rice genomic  
 CC library. Four positive clones were identified and sequenced. The  
 CC promoter region (tag = b) corresponds to nucleotides -785 to  
 CC plus 59 using conventional nucleotide numbering, i.e. where  
 CC transcription start site is plus 1.  
 XX  
 SQ Sequence 2075 BP; 481 A; 610 C; 502 G; 482 T; 0 other;

Query Match 7.2%; Score 30.4; DB 13; Length 2075;  
 Best Local Similarity 47.0%; Pred. No. 11;  
 Matches 94; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

Qy 32 cgagatcatgaggaagcaataactcccctgtgctctacttctgcctgatggcccttgctgt 91  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 1490 caaggtgaaggagatcaagaaaggccgcctcgccatgttctccatgttcggcttcttctgt 1549  
 Qy 92 agctgctatggtctgtgtcatgtacaccacctcggcacaaagcaggaaggagtggctacaa 151  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 1550 ccaggccatcgctaccgggaagggccccctcgagaacctcgccgaccacctcgccgaccc 1609  
 Qy 152 ctcgtagaacctgatggaaggggtggatacaactctgttcccatcaacggcggtggcag 211  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 1610 cgtcaacaacaacgcctgggctgacgccaccaacttcgtccccggcaagtgaagtggggg 1669  
 Qy 212 cccctagctaggcggtggat 231  
 | | | | | | | | | | | | | | | |

Db 1670 accgtagcttagcagtggtt 1689

RESULT 6

AAF74625

ID AAF74625 standard; cDNA; 3326 BP.

XX

AC AAF74625;

XX

DT 14-MAY-2001 (first entry)

XX

DE Human GLI-1 nucleotide sequence SEQ ID NO:27.

XX

KW SUFUH; GLI-1; Sonic hedgehog-patched signalling pathway; cancer;  
KW cell differentiation; tissue development; ss.

XX

OS Homo sapiens.

XX

PN WO200112655-A1.

XX

PD 22-FEB-2001.

XX

PF 14-AUG-2000; 2000WO-SE01576.

XX

PR 13-AUG-1999; 99SE-0002899.

XX

PA (KARO-) KAROLINSKA INNOVATIONS AB.

XX

PI Toftgard R;

XX

DR WPI; 2001-211199/21.

XX

PT Novel peptides comprising fragments of two components of sonic  
PT hedgehog-patched signaling pathway, GLI-1 and SUFUH, useful for  
PT treating cancer and diseases influencing cell differentiation and  
PT tissue development -

XX

PS Example 3; Page 111-112; 115pp; English.

XX

CC The present invention describes peptides consisting of fragments of GLI-1  
CC and SUFUH, respectively which are able to specifically bind to SUFUH and  
CC GLI-1, respectively. GLI-1 and SUFUH are components which interact in the  
CC Sonic hedgehog (Shh)-patched (Ptch) signalling pathway. The present  
CC invention also describes: (1) DNA sequences encoding the peptides; and  
CC (2) a monoclonal antibody or an antibody fragment directed against the  
CC peptides. The peptides have cytostatic activity, and can be used as  
CC Shh-Ptch signalling pathway modulators. The peptides and monoclonal  
CC antibodies against them can be used for preparing a pharmaceutical  
CC composition for treating cancer. The peptides on contact with the GLI-1  
CC and SUFUH in vivo affects the Shh-Ptch signalling pathway which is used  
CC in the treatment of cancer. The peptides comprising the peptide fragments  
CC of the signalling pathway are also useful for treating other diseases  
CC influencing cell differentiation and tissue development. The present  
CC sequence represents the human GLI-1 nucleotide sequence, which is used  
CC in an example from the present invention.

XX

SQ Sequence 3326 BP; 715 A; 1087 C; 879 G; 645 T; 0 other;

Query Match 7.1%; Score 30.2; DB 22; Length 3326;  
Best Local Similarity 53.9%; Pred. No. 16;  
Matches 62; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy      72  tgcctgatggcccttgctcgtagctgctatgggtctgtgtcatgtacaccacctcggcacia 131
          ||  || || || || || |  || || || || || |  |
Db    1442  tggacgaggggaccttgcatgtgctggcactgggtctgtccactcttcgccgccttgagaacc 1501

Qy     132  gcaggaaggagtggtacaaactcgtaacgaacctgatggaaggggtggatacaact 186
          || || || || || || |  || || |  || || || || ||
Db    1502  tcaggctgggaccagctacatcaactccggccaatagggacccgggtctcaaact 1556
```

RESULT 7

AAD12302

ID AAD12302 standard; cDNA; 3600 BP.

XX

AC AAD12302;

XX

DT 16-OCT-2001 (first entry)

XX

DE Human Cubitus interruptus (Ci) homologue, GLI-1 cDNA.

XX

KW Human; transgenic non-human animal; Cubitus interruptus; Ci; GLI-1;

KW basal cell carcinoma; BCC model system; tumour; screening;

KW anti-cancer; trichoepithelioma; cylindroma; trichoblastoma; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 79..3399

FT /\*tag= a

FT /product= "Human Ci homologue, GLI-1"

XX

PN WO200156376-A1.

XX

PD 09-AUG-2001.

XX

PF 02-FEB-2001; 2001WO-SE00204.

XX

PR 03-FEB-2000; 2000SE-0000345.

XX

PA (KARO-) KAROLINSKA INNOVATIONS AB.

XX

PI Toftgard R;

XX

DR WPI; 2001-488828/53.

DR P-PSDB; AAE06644.

XX

PT Transgenic non-human animal useful as basal cell carcinoma model system

PT to identify anti-cancer drug candidates, overexpresses transgene

PT encoding GLI-1 protein which is a human homolog to Cubitus interruptus

PT -

XX

PS Claim 6; Page 25-26; 33pp; English.



PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.

PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.



PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.

PR 29-OCT-1999; 99US-0162142.

Query Match 7.1%; Score 30; DB 21; Length 5118;  
Best Local Similarity 55.9%; Pred. No. 22;  
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```
Qy 43 ggaagcaatactcccctgtgctctacttctgcctgatggcccttgctcgtagctgctatgg 102
      ||||| | || || ||| | || | || | ||| | | ||| ||| ||
Db 2334 ggaagctaagcttctctgatgcccgacctttgattaatgtctgtgatcgttttggctttgt 2393

Qy 103 tctgtgtcatgtacaccacctcggcacaagcaggaaggagtg 144
      | || | || |||| |||| || | | | |
Db 2394 acctgatcttactcattacctctacacaaacaacatgctgcg 2435
```

RESULT 9

AAQ14640

ID AAQ14640 standard; DNA; 1229 BP.

XX

AC AAQ14640;

XX

DT 30-JAN-1992 (first entry)

XX

DE Plasmid pGB18ARR insert encoding a human temporal lobe PDE.

XX

KW brain; pRATDPD; cAMP; phosphodiesterase; complemetation analysis; ss.

XX

OS Homo sapiens.

XX

PN WO9116457-A.

XX

PD 31-OCT-1991.

XX

PF 19-APR-1991; 91WO-US02714.

XX

PR 20-APR-1990; 90US-0511715.

XX

PA (COLD-) COLD SPRING HARBOR.

XX

PI Wigler MH, Colicelli JJ;

XX

DR WPI; 1991-339841/46.

XX

PT Complementary screening for genes and prods. - e.g. RAS protein

PT and cAMP, that modify, complement or suppress genetic defect and

PT correct associated phenotypic alteration

XX

PS Example 2; Page 124; 169pp; English.

XX

CC Plasmid pRATDPD was isolated from a rat brain cDNA library. It is  
CC thought to encode a cyclic nucleotide PDE. The RATDPD cDNA was used  
CC as a probe to isolate plasmid pGB18ARR by low stringency screens of  
CC a human temporal lobe cDNA library. The inventors have classified  
CC GB18ARR in cAMP-specific PDE class IV2 along with TM3 (AAQ14630).

XX

SQ Sequence 1229 BP; 281 A; 360 C; 341 G; 247 T; 0 other;

Query Match 7.0%; Score 29.8; DB 12; Length 1229;  
 Best Local Similarity 51.9%; Pred. No. 14;  
 Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

Qy 281 cacactgccgctccccacgctaaatttggggggtacagtgcacacgctagccgattaac 340
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 846 cacatccgcactcccagctcctggtggcggggggtcaggtggagaccctacctgatcccc 905

Qy 341 ggctcacgctaccagggcgtctacgcggatgtgccccctagccagcttctctctccccct 400
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 906 agacctctgtccctgttccccctccactcctccccctcactccccctgtcccccgaccacct 965

Qy 401 cgttctgtg 409
      | | | | |
Db 966 cctcctctg 974
  
```

RESULT 10

AAT34381

ID AAT34381 standard; cDNA; 1230 BP.

XX

AC AAT34381;

XX

DT 10-OCT-1996 (first entry)

XX

DE Plasmid pGB18ARR insert.

XX

KW Human; glioblastoma cell; plasmid; mammalian; complementation; pPDE2RR;  
 KW probe; yeast; pPDET; pPDE10X inv; temporal lobe; cDNA library; pRATPDP;  
 KW pTM72; pGB14; pGB18ARR; pTM3; pJC44x; pGB25; phosphodiesterase family IV;  
 KW pPDE18; pPDE21; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 3..1157

FT /\*tag= a

FT /product= cAMP phosphodiesterase

XX

PN US5527896-A.

XX

PD 18-JUN-1996.

XX

PF 20-APR-1990; 90US-0511715.

XX

PR 19-APR-1991; 91US-0688352.

PR 20-APR-1990; 90US-0511715.

XX

PA (COLD-) COLD SPRING HARBOR LAB.

XX

PI Colicelli JJ, Wigler MH;

XX

DR WPI; 1996-299902/30.

DR P-PSDB; AAW00097.

XX

PT DNA mols. isolated from human glioblastoma cells - encode

PT RAS-related or cyclic nucleotide phosphodiesterase proteins  
XX  
PS Claim 5; Column 125-128; 101pp; English.  
XX  
CC The sequences given in AAT34377-84 represent plasmid fragments which  
CC were isolated by hybridisation with mammalian genes cloned by  
CC complementation. These sequences were isolated using probes derived  
CC from the sequences given in AAT34366-76 which were cloned via  
CC complementation in yeast. Plasmids pPDET, pPDE10X inv and pPDE2RR were  
CC isolated by low stringency hybridisation screens of a human temporal  
CC lobe cDNA library using the pRATPDP insert as a probe. Comparison of  
CC the nucleotide sequences given in AAT34377-79 indicated that the inserts  
CC are representatives of the same genetic locus as the insert in pTM72.  
CC Plasmids pGB14 and pGB18ARR were obtained in the same manner. DNA  
CC sequence analysis revealed that they are representatives of the same  
CC genetic locus as the inserts in pTM3 and pJC44x. Plasmid pGB25 was also  
CC obtained at low stringency hybridisation using the pRATDPD insert as a  
CC probe. Judged by its nucleotide and deduced amino acid sequence it  
CC represents a novel member of the phosphodiesterase family IV. The cDNA  
CC insert of pGB25 was used as a probe to obtain pPDE18 and pPDE21. The  
CC cDNA of pPDE18 represents the same locus as that of pGB25 and contains  
CC more sequence information than the pGB25 cDNA. The pPDE21 insert  
CC represents a fourth member of phosphodiesterase family IV. The  
CC assignment to family IV is based solely on sequence relationships.  
XX  
SQ Sequence 1230 BP; 281 A; 360 C; 341 G; 247 T; 1 other;

Query Match 7.0%; Score 29.8; DB 17; Length 1230;  
Best Local Similarity 51.9%; Pred. No. 14;  
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

Qy 281 cacactgccgctccccacgctaaatttgggggctaagtcacacgctagccgattaac 340
      |||| | |||| | | |||| | || | || || | || |
Db 847 cacatccgcactcccagctcctggtggcggggggtcaggtggagaccctacctgatcccc 906

Qy 341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
      | | | | | | | | | | | | | | | | | | | |
Db 907 agacctctgtccctgttccccctccactcctccccctcactccccctgctcccccgaccacct 966

Qy 401 cgttctgtg 409
      | | | | |
Db 967 cctcctctg 975

```

RESULT 11  
AAZ32251  
ID AAZ32251 standard; cDNA; 1230 BP.  
XX  
AC AAZ32251;  
XX  
DT 19-JAN-2000 (first entry)  
XX  
DE Human temporal lobe phosphodiesterase pGB18ARR encoding cDNA.  
XX  
KW Phosphodiesterase; dunce-like phosphodiesterase; PDE; DPD; cAMP;  
KW RAS-related protein; immunoreactive; detection; genetic defect;

KW bronchodilation; increased myocardial contractility;  
XX anti-inflammation; ss.  
OS Homo sapiens.  
XX  
PN US5977305-A.  
XX  
PD 02-NOV-1999.  
XX  
PF 07-JUN-1995; 95US-0474379.  
XX  
PR 01-MAR-1994; 94US-0206188.  
PR 20-APR-1990; 90US-0511715.  
PR 19-APR-1991; 91US-0688352.  
XX  
PA (COLD-) COLD SPRING HARBOR LAB.  
XX  
PI Colicelli JJ, Wigler MH;  
XX  
DR WPI; 1999-619709/53.  
DR P-PSDB; AAY49817.  
XX  
PT New isolated RAS-related polypeptides and mammalian cyclic nucleotide  
PT phosphodiesterases, used for screening for agents which can modify  
PT complement or suppress genetic defects -  
XX  
PS Example 2; Column 133-138; 145pp; English.  
XX  
CC The present invention describes new isolated RAS-related polypeptides  
CC and mammalian cyclic nucleotide phosphodiesterases (PDEs). RAS-related  
CC polypeptides are capable of complementing a defective RAS function in  
CC yeast. The products can be used for screening for agents which can  
CC modify, complement or suppress a genetic defect in a biochemical  
CC pathway in which cAMP participates, or in a biochemical pathway which  
CC is controlled, directly or indirectly, by a RAS protein and other  
CC proteins affecting cell growth and maintenance. Developing agents that  
CC will selectively act upon PDEs is directed toward reproducing the  
CC desirable effects of cyclic nucleotides, e.g. bronchodilation,  
CC increased myocardial contractility, anti-inflammation, yet without  
CC causing the undesirable effects, e.g. increased heart rate or enhanced  
CC lipolysis. The products can also be used for therapeutic, diagnostic  
CC and prognostic uses. AAZ32229 to AAZ32285, and AAY49803 to AAY49830,  
CC represent sequences used in the exemplification of the present  
CC invention.  
XX  
SQ Sequence 1230 BP; 281 A; 360 C; 341 G; 247 T; 1 other;

Qy	281	cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac	340
Db	847	cacatccgcactcccagctcctgggtggcggggggtcaggtggagaccctacctgatcccc	906
Qy	341	ggctcacgctaccaggcgctctacgcggatgtgcccccctagccagcttctctctccccct	400

```

          | | | | | | | | | | | | | | | | | | | |
Db      907 agacctctgtccctgttccccctccactcctccccctcactccccctgctcccccgaccacct 966

Qy      401 cgttctgtg 409
          | | | | |
Db      967 cctcctctg 975

```

RESULT 12

AAA88186

ID AAA88186 standard; cDNA; 1230 BP.

XX

AC AAA88186;

XX

DT 14-DEC-2000 (first entry)

XX

DE pGB18ARR human temporal lobe insert nucleotide sequence SEQ ID NO:39.

XX

KW Detection; mammalian gene; yeast; microorganism; identification;

KW phenotype; characteristic; dunce-like phosphodiesterase; PDE; RAS;

KW RAS-related protein; genetic defect; hybridisation; probe; ss.

XX

OS Homo sapiens.

OS Synthetic.

XX

PN US6100025-A.

XX

PD 08-AUG-2000.

XX

PF 01-MAR-1994; 94US-0206188.

XX

PR 20-APR-1990; 90US-0511715.

PR 19-APR-1991; 91US-0688352.

XX

PA (COLD-) COLD SPRING HARBOR LAB.

XX

PI Colicelli JJ, Wigler MH;

XX

DR WPI; 2000-531664/48.

DR P-PSDB; AAB20628.

XX

PT Novel isolated DNA encoding a mammalian cyclic nucleotide

PT phosphodiesterase is present in plasmids pPDE46, pPDE43 or pPDE339 and

PT is used to modify a genetic defect in a biochemical pathway in which

PT cAMP participates -

XX

PS Example 2; Column 139-144; 145pp; English.

XX

CC The present invention describes a purified and isolated DNA (I) which  
 CC encodes a mammalian cyclic nucleotide phosphodiesterase and is an insert  
 CC present in the plasmids pPDE46 (ATCC 69552), pPDE43 (ATCC 69551) or  
 CC pPDE339 (ATCC 69550). The DNA molecules are used to modify, complement  
 CC or suppress a genetic defect in a biochemical pathway in which cAMP  
 CC participates and are also used as hybridisation probes. The present  
 CC invention also describes methods for detecting mammalian genes encoding  
 CC proteins which can function in microorganisms, particularly yeast, to  
 CC modify, complement, or suppress a genetic defect associated with an

CC identifiable phenotypic alteration or characteristic in the  
CC microorganism. AAA88162 to AAA88218 and AAB29614 to AAB20640 represent  
CC sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 1230 BP; 281 A; 360 C; 341 G; 247 T; 1 other;

Query Match 7.0%; Score 29.8; DB 21; Length 1230;  
Best Local Similarity 51.9%; Pred. No. 14;  
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
Qy 281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340
      |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 847 cacatccgcactcccagctcctggtggcggggggtcaggtggagaccctacctgatcccc 906

Qy 341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 907 agacctctgtccctgttccccctccactcctccccctcactccccctgctcccccgaccacct 966

Qy 401 cgttctgtg 409
      | | | | |
Db 967 cctcctctg 975
```

RESULT 13

AAZ32285

ID AAZ32285 standard; cDNA; 1481 BP.

XX

AC AAZ32285;

XX

DT 19-JAN-2000 (first entry)

XX

DE Nucleotide sequence SEQ ID NO:87.

XX

KW Phosphodiesterase; dunce-like phosphodiesterase; PDE; DPD; cAMP;

KW RAS-related protein; immunoreactive; detection; genetic defect;

KW bronchodilation; increased myocardial contractility;

KW anti-inflammation; ss.

XX

OS Unidentified.

XX

PN US5977305-A.

XX

PD 02-NOV-1999.

XX

PF 07-JUN-1995; 95US-0474379.

XX

PR 01-MAR-1994; 94US-0206188.

PR 20-APR-1990; 90US-0511715.

PR 19-APR-1991; 91US-0688352.

XX

PA (COLD-) COLD SPRING HARBOR LAB.

XX

PI Colicelli JJ, Wigler MH;

XX

DR WPI; 1999-619709/53.

DR P-PSDB; AAY49830.

XX  
PT New isolated RAS-related polypeptides and mammalian cyclic nucleotide  
PT phosphodiesterases, used for screening for agents which can modify  
PT complement or suppress genetic defects -  
XX  
PS Disclosure; Column 211-214; 145pp; English.  
XX  
CC The present invention describes new isolated RAS-related polypeptides  
CC and mammalian cyclic nucleotide phosphodiesterases (PDEs). RAS-related  
CC polypeptides are capable of complementing a defective RAS function in  
CC yeast. The products can be used for screening for agents which can  
CC modify, complement or suppress a genetic defect in a biochemical  
CC pathway in which cAMP participates, or in a biochemical pathway which  
CC is controlled, directly or indirectly, by a RAS protein and other  
CC proteins affecting cell growth and maintenance. Developing agents that  
CC will selectively act upon PDEs is directed toward reproducing the  
CC desirable effects of cyclic nucleotides, e.g. bronchodilation,  
CC increased myocardial contractility, anti-inflammation, yet without  
CC causing the undesirable effects, e.g. increased heart rate or enhanced  
CC lipolysis. The products can also be used for therapeutic, diagnostic  
CC and prognostic uses. AAZ32229 to AAZ32285, and AAY49803 to AAY49830,  
CC represent sequences used in the exemplification of the present  
CC invention.  
XX  
SQ Sequence 1481 BP; 349 A; 430 C; 389 G; 313 T; 0 other;

Query Match 7.0%; Score 29.8; DB 20; Length 1481;  
Best Local Similarity 51.9%; Pred. No. 15;  
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340  
| | | | | | | | | | | | | | | | | | | | | |  
Db 957 cacatccgcactcccagctcctggtggcggggggtcaggtggagaccctacctgatcccc 1016  
Qy 341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1017 agacctctgtccctgttccccctccactcctccccctcactccccctgctcccccgaccacct 1076  
Qy 401 cgttctgtg 409  
| | | | |  
Db 1077 cctcctctg 1085

RESULT 14  
AAT34373  
ID AAT34373 standard; cDNA; 2702 BP.  
XX  
AC AAT34373;  
XX  
DT 09-OCT-1996 (first entry)  
XX  
DE Plasmid pJC44x (ATCC 68603) insert.  
XX  
KW Human; glioblastoma; RAS-related protein; cell line U118MG; pJC265;  
KW yeast expression vector; pADNS; pADANS; fusion protein; rat pRATDPD;  
KW alcohol dehydrogenase protein; heat shock sensitivity; S. cerevisiae;



KW TK161-R2V; pJC44x; pJC99; pJC310; cAMP phosphodiesterase; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..2701  
 FT /\*tag= a  
 FT /product= cAMP phosphodiesterase  
 XX  
 PN US5527896-A.  
 XX  
 PD 18-JUN-1996.  
 XX  
 PF 20-APR-1990; 90US-0511715.  
 XX  
 PR 19-APR-1991; 91US-0688352.  
 PR 20-APR-1990; 90US-0511715.  
 XX  
 PA (COLD-) COLD SPRING HARBOR LAB.  
 XX  
 PI Colicelli JJ, Wigler MH;  
 XX  
 DR WPI; 1996-299902/30.  
 DR P-PSDB; AAW00091.  
 XX  
 PT DNA mols. isolated from human glioblastoma cells - encode  
 PT RAS-related or cyclic nucleotide phosphodiesterase proteins  
 XX  
 PS Claim 4; Column 39-44; 101pp; English.  
 XX  
 CC This sequence represents a plasmid fragment which contains a human  
 CC glioblastoma cell cDNA inserts encoding a cAMP phosphodiesterase. The  
 CC cDNA was derived from the human glioblastoma cell line U118MG and  
 CC transferred into two yeast expression vectors, pADNS and pADANS.  
 CC Plasmid pADANS differs from pADNS in that the mRNA transcribed will  
 CC direct the synthesis of a fusion protein including an N-terminal  
 CC portion derived from the alcohol dehydrogenase protein, and the  
 CC remainder from the mammalian cDNA insert. The two cDNA expression  
 CC libraries created were screen for cDNA's capable of correcting the  
 CC heat shock sensitivity of the S. cerevisiae host TK161-R2V. Four  
 CC different inserts contained in plasmids pJC44x, pJC99, pJC265 and  
 CC pJC310 (see also AAT34366-68) were discovered. The insert of pJC44x  
 CC was shown to be homologous to the rat pRATDPD gene and biochemical  
 CC analysis of the cellular lysates demonstrated that it encoded a cAMP  
 CC phosphodiesterase. The inserts of pJC99, pJC265 and pJC310 showed no  
 CC significant homology to previous isolated genes.  
 XX  
 SQ Sequence 2702 BP; 574 A; 887 C; 757 G; 484 T; 0 other;

Query Match 7.0%; Score 29.8; DB 17; Length 2702;  
 Best Local Similarity 51.9%; Pred. No. 20;  
 Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340  
 |||| | |||| | | |||| | || | || | || |  
 Db 2534 cacatccgcactcccagctcctggtggcggggggtcaggtggagaccctacctgatcccc 2593

Qy 341 ggctcacgctaccagggcgctctacgcggatgtgccccctagccagcttctctctccccct 400  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 2594 agacctctgtccctgttccccctccactcctccccctcactccccctgctcccccgaccacct 2653

Qy 401 cgttctgtg 409  
 | | | | |  
 Db 2654 cctcctctg 2662

RESULT 15

AAZ32236

ID AAZ32236 standard; cDNA; 2702 BP.

XX

AC AAZ32236;

XX

DT 19-JAN-2000 (first entry)

XX

DE Human glioblastoma cell cAMP phosphodiesterase pJC44x encoding cDNA.

XX

KW Phosphodiesterase; dunce-like phosphodiesterase; PDE; DPD; cAMP;

KW RAS-related protein; immunoreactive; detection; genetic defect;

KW bronchodilation; increased myocardial contractility;

KW anti-inflammation; ss.

XX

OS Homo sapiens.

XX

PN US5977305-A.

XX

PD 02-NOV-1999.

XX

PF 07-JUN-1995; 95US-0474379.

XX

PR 01-MAR-1994; 94US-0206188.

PR 20-APR-1990; 90US-0511715.

PR 19-APR-1991; 91US-0688352.

XX

PA (COLD-) COLD SPRING HARBOR LAB.

XX

PI Colicelli JJ, Wigler MH;

XX

DR WPI; 1999-619709/53.

DR P-PSDB; AAY49804.

XX

PT New isolated RAS-related polypeptides and mammalian cyclic nucleotide

PT phosphodiesterases, used for screening for agents which can modify

PT complement or suppress genetic defects -

XX

PS Example 1; Column 53-59; 145pp; English.

XX

CC The present invention describes new isolated RAS-related polypeptides

CC and mammalian cyclic nucleotide phosphodiesterases (PDEs). RAS-related

CC polypeptides are capable of complementing a defective RAS function in

CC yeast. The products can be used for screening for agents which can

CC modify, complement or suppress a genetic defect in a biochemical

CC pathway in which cAMP participates, or in a biochemical pathway which

CC is controlled, directly or indirectly, by a RAS protein and other

CC proteins affecting cell growth and maintenance. Developing agents that  
CC will selectively act upon PDEs is directed toward reproducing the  
CC desirable effects of cyclic nucleotides, e.g. bronchodilation,  
CC increased myocardial contractility, anti-inflammation, yet without  
CC causing the undesirable effects, e.g. increased heart rate or enhanced  
CC lipolysis. The products can also be used for therapeutic, diagnostic  
CC and prognostic uses. AAZ32229 to AAZ32285, and AAY49803 to AAY49830,  
CC represent sequences used in the exemplification of the present  
CC invention.

XX

SQ Sequence 2702 BP; 574 A; 887 C; 757 G; 484 T; 0 other;

Query Match 7.0%; Score 29.8; DB 20; Length 2702;  
Best Local Similarity 51.9%; Pred..No. 20;  
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 281 cacactgccgctcccccacgctaaatttgggggctacagtgcacacgctagcogattaac 340  
| | | | | | | | | | | | | | | | | | | | | |  
Db 2534 cacatccgcactcccagctcctggtggcggggggtcaggtggagaccctacctgatcccc 2593  
  
Qy 341 ggctcacgctaccaggcgctctacgggatgtgccccctagccagcttctctctccccct 400  
| | | | | | | | | | | | | | | | | | | | | |  
Db 2594 agacctctgtccctgttccctccactcctccctcactccctgctcccccgaccacct 2653  
  
Qy 401 cgttctgtg 409  
| | | | |  
Db 2654 cctcctctg 2662

Search completed: February 7, 2002, 11:00:01  
Job time: 4987 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 10:51:46 ; Search time 172.96 Seconds  
(without alignments)  
556.505 Million cell updates/sec

Title: US-09-394-745-6332  
Perfect score: 425  
Sequence: 1 cggacgcgtgggtgcaattt.....tgtggtgcctctctcaacct 425

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	29.8	7.0	1229	5	PCT-US91-02714-38	Sequence 38, Appl
2	29.8	7.0	1230	1	US-07-688-352C-39	Sequence 39, Appl
3	29.8	7.0	1230	2	US-08-474-379C-39	Sequence 39, Appl
4	29.8	7.0	1230	3	US-09-146-249A-39	Sequence 39, Appl
5	29.8	7.0	1230	3	US-08-206-188B-39	Sequence 39, Appl
6	29.8	7.0	1481	2	US-08-474-379C-87	Sequence 87, Appl
7	29.8	7.0	2702	1	US-07-688-352C-11	Sequence 11, Appl
8	29.8	7.0	2702	2	US-08-474-379C-11	Sequence 11, Appl
9	29.8	7.0	2702	3	US-09-146-249A-11	Sequence 11, Appl
10	29.8	7.0	2702	3	US-08-206-188B-11	Sequence 11, Appl
11	29.8	7.0	3131	1	US-07-688-352C-21	Sequence 21, Appl
12	29.8	7.0	3131	3	US-09-146-249A-21	Sequence 21, Appl
13	29.8	7.0	3131	3	US-08-206-188B-21	Sequence 21, Appl
14	29.8	7.0	3131	5	PCT-US91-02714-20	Sequence 20, Appl
15	29.8	7.0	3705	2	US-08-474-379C-64	Sequence 64, Appl
16	29.8	7.0	3705	3	US-09-146-249A-64	Sequence 64, Appl
17	29.8	7.0	3705	3	US-08-206-188B-64	Sequence 64, Appl
18	29.6	7.0	740	4	US-09-020-956-17	Sequence 17, Appl
19	29.6	7.0	740	4	US-09-030-607-17	Sequence 17, Appl
20	28.8	6.8	4550	4	US-09-103-663-35	Sequence 35, Appl
21	28.6	6.7	498	4	US-09-037-990B-6	Sequence 6, Appli
22	28.6	6.7	556	4	US-09-037-990B-7	Sequence 7, Appli
23	28.4	6.7	986	2	US-08-713-825-2	Sequence 2, Appli
24	28.4	6.7	986	3	US-09-199-842-2	Sequence 2, Appli
25	28	6.6	1897	1	US-08-453-477-1	Sequence 1, Appli
26	28	6.6	1897	1	US-08-453-461-1	Sequence 1, Appli
27	28	6.6	7218	1	US-08-232-463-14	Sequence 14, Appl
c 28	27.6	6.5	751	4	US-09-020-956-12	Sequence 12, Appl
c 29	27.6	6.5	751	4	US-09-030-607-12	Sequence 12, Appl
30	27.6	6.5	3530	3	US-08-704-711A-10	Sequence 10, Appl
c 31	27.4	6.4	801	4	US-09-020-956-16	Sequence 16, Appl
c 32	27.4	6.4	801	4	US-09-030-607-16	Sequence 16, Appl
33	27.4	6.4	933	3	US-08-808-148-2	Sequence 2, Appli
34	27.4	6.4	1289	4	US-09-020-956-111	Sequence 111, App
35	27.4	6.4	1289	4	US-09-030-607-111	Sequence 111, App

	36	27.4	6.4	16075	3	US-09-096-942-1	Sequence 1, Appli
	37	27.4	6.4	16075	3	US-09-096-867-1	Sequence 1, Appli
c	38	27.2	6.4	1557	3	US-09-329-418-2	Sequence 2, Appli
c	39	27.2	6.4	1557	4	US-09-531-914-2	Sequence 2, Appli
c	40	27.2	6.4	1873	3	US-09-329-418-1	Sequence 1, Appli
c	41	27.2	6.4	1873	4	US-09-531-914-1	Sequence 1, Appli
c	42	27.2	6.4	2870	1	US-08-468-036-28	Sequence 28, Appl
c	43	27.2	6.4	2870	2	US-08-376-843-28	Sequence 28, Appl
c	44	27.2	6.4	15328	2	US-08-888-497-33	Sequence 33, Appl
c	45	27.2	6.4	15328	5	PCT-US94-07926-33	Sequence 33, Appl

# ALIGNMENTS

```

RESULT      1
PCT-US91-02714-38
; Sequence 38, Application PC/TUS9102714
;  GENERAL INFORMATION:
;    APPLICANT:  Wigler, Michael H.
;    APPLICANT:  Colicelli, John J.
;    TITLE OF INVENTION:  Cloning by Complementation and Related
;    TITLE OF INVENTION:  Processes
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:    Two First National Plaza, 20 South Clark
;      STREET:    Street
;      CITY:      Chicago
;      STATE:     Illinois
;      COUNTRY:   USA
;      ZIP:       60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:     IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:     PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/02714
;      FILING DATE:       19910419
;      CLASSIFICATION:    435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/511,715
;      FILING DATE:       20-APR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Borun, Michael F.
;      REGISTRATION NUMBER:  25447
;      REFERENCE/DOCKET NUMBER:  27805/30197
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 346-5750
;      TELEFAX:   (312) 984-9740
;      TELEX:     25-3856
;    INFORMATION FOR SEQ ID NO:  38:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  1229 base pairs
;      TYPE:    NUCLEIC ACID

```

; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
PCT-US91-02714-38

Query Match 7.0%; Score 29.8; DB 5; Length 1229;  
Best Local Similarity 51.9%; Pred. No. 2.6;  
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
Qy 281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340
    |||| | |||| | |||| | |||| | |||| | |||| |
Db 846 CACATCCGCACTCCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 905

Qy 341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
    | | | | | | | | | | | | | | | | | | | | | |
Db 906 AGACCTCTGTCCCTGTTCCCTCCACTCCTCCCTCACTCCCTGCTCCCCCGACCACCT 965

Qy 401 cgttctgtg 409
    | | | | |
Db 966 CCTCCTCTG 974
```

RESULT 2

US-07-688-352C-39

; Sequence 39, Application US/07688352C

; Patent No. 5527896

; GENERAL INFORMATION:

; APPLICANT: Wigler, Michael H.

; APPLICANT: Colicelli, John J.

; TITLE OF INVENTION: Cloning by Complementation and Related

; TITLE OF INVENTION: Processes

; NUMBER OF SEQUENCES: 57

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

; ADDRESSEE: Bicknell

; STREET: Two First National Plaza, 20 South Clark

; STREET: Street

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60603

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/688,352C

; FILING DATE: 19910419

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/511,715

; FILING DATE: 20-APR-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Borun, Michael F.

; REGISTRATION NUMBER: 25447

```

; REFERENCE/DOCKET NUMBER: 27805/30197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1230 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..1156
US-07-688-352C-39

```

```

Query Match          7.0%; Score 29.8; DB 1; Length 1230;
Best Local Similarity 51.9%; Pred. No. 2.6;
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

```

Qy  281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340
      ||||  | ||||  ||  ||||  |  |||  || ||||  |||  |
Db  847 CACATCCGCACTCCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 906

Qy  341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
      |  |  |  |||  |  |||  |  |  |||  || ||||  |  || |||
Db  907 AGACCTCTGTCCCTGTTCCCTCCACTCCTCCCTCACTCCCTGCTCCCCCGACCACCT 966

Qy  401 cgttctgtg 409
      |  ||||
Db  967 CCTCCTCTG 975

```

```

RESULT 3
US-08-474-379C-39
; Sequence 39, Application US/08474379C
; Patent No. 5977305
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
; TITLE OF INVENTION: PROCESSES
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,379C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/206,188
; FILING DATE: 01-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/688,352
; FILING DATE: 19-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 27866/32771
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1230 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..1154
US-08-474-379C-39

```

```

Query Match          7.0%; Score 29.8; DB 2; Length 1230;
Best Local Similarity 51.9%; Pred. No. 2.6;
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

```

Qy 281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340
    |||| | |||| | | |||| | || | || || | || |
Db 847 CACATCCGCACTCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 906

Qy 341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
    | | | | || | | || | | | | | | | | | | | |
Db 907 AGACCTCTGTCCCTGTTCCCTCCACTCCTCCCTCACTCCCTGCTCCCCCGACCACCT 966

Qy 401 cgttctgtg 409
    | | || |
Db 967 CCTCCTCTG 975

```

```

RESULT 4
US-09-146-249A-39
; Sequence 39, Application US/09146249A
; Patent No. 6069240
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.

```



```

; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,249A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1230 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..1156
US-09-146-249A-39

```

```

Query Match          7.0%; Score 29.8; DB 3; Length 1230;
Best Local Similarity 51.9%; Pred. No. 2.6;
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

```

Qy  281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340
      |||| | |||| | | |||| | || | || | || | || |
Db  847 CACATCCGCACTCCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 906

Qy  341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
      | | | | | | | | | | | | | | | | | | | | | |
Db  907 AGACCTCTGTCCCTGTTCCCTCCACTCCTCCCTCACTCCCCTGCTCCCCCGACCACCT 966

Qy  401 cgttctgtg 409
      | | | | |
Db  967 CCTCCTCTG 975

```

RESULT 5  
 US-08-206-188B-39  
 ; Sequence 39, Application US/08206188B  
 ; Patent No. 6100025  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wigler, Michael H.  
 ; APPLICANT: Colicelli, John J.  
 ; TITLE OF INVENTION: Cloning by Complementation and Related  
 ; TITLE OF INVENTION: Processes  
 ; NUMBER OF SEQUENCES: 84  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/206,188B  
 ; FILING DATE: 01-MAR-1994  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/511,715  
 ; FILING DATE: 20-APR-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Clough, David W.  
 ; REGISTRATION NUMBER: 36107  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312-474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 39:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1230 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 3..1156  
 US-08-206-188B-39

Query Match 7.0%; Score 29.8; DB 3; Length 1230;  
 Best Local Similarity 51.9%; Pred. No. 2.6;  
 Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340  
 |||| | |||| | | |||| | || | || | || |

Db 847 CACATCCGCACTCCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 906  
 Qy 341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 907 AGACCTCTGTCCCTGTTCCCCTCCACTCCTCCCCTCACTCCCCTGCTCCCCCGACCACCT 966  
 Qy 401 cgttctgtg 409  
 | | | | |  
 Db 967 CCTCCTCTG 975

RESULT 6  
 US-08-474-379C-87  
 ; Sequence 87, Application US/08474379C  
 ; Patent No. 5977305  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wigler, Michael H.  
 ; APPLICANT: Colicelli, John J.  
 ; TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED  
 ; TITLE OF INVENTION: PROCESSES  
 ; NUMBER OF SEQUENCES: 88  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 233 South Wacker Drive/6300 Sears Tower  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/474,379C  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/511,715  
 ; FILING DATE: 20-APR-1990  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/206,188  
 ; FILING DATE: 01-MAR-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/688,352  
 ; FILING DATE: 19-APR-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Clough, David W.  
 ; REGISTRATION NUMBER: 36,107  
 ; REFERENCE/DOCKET NUMBER: 27866/32771  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (312) 474-6300  
 ; TELEFAX: (312) 474-0448  
 ; INFORMATION FOR SEQ ID NO: 87:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1481 base pairs  
 ; TYPE: nucleic acid

; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1008  
US-08-474-379C-87

Query Match 7.0%; Score 29.8; DB 2; Length 1481;  
Best Local Similarity 51.9%; Pred. No. 2.8;  
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340  
| | | | | | | | | | | | | | | | | | | | | |  
Db 957 CACATCCGCACTCCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 1016  
  
Qy 341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1017 AGACCTCTGTCCCTGTTCCCCTCCACTCCTCCCCTCACTCCCCTGCTCCCCCGACCACCT 1076  
  
Qy 401 cgttctgtg 409  
| | | | |  
Db 1077 CTCCTCTG 1085

RESULT 7  
US-07-688-352C-11  
; Sequence 11, Application US/07688352C  
; Patent No. 5527896  
; GENERAL INFORMATION:  
; APPLICANT: Wigler, Michael H.  
; APPLICANT: Colicelli, John J.  
; TITLE OF INVENTION: Cloning by Complementation and Related  
; TITLE OF INVENTION: Processes  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
; ADDRESSEE: Bicknell  
; STREET: Two First National Plaza, 20 South Clark  
; STREET: Street  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/688,352C  
; FILING DATE: 19910419  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/511,715  
; FILING DATE: 20-APR-1990

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Borun, Michael F.
; REGISTRATION NUMBER: 25447
; REFERENCE/DOCKET NUMBER: 27805/30197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2702 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..2701
US-07-688-352C-11

```

```

Query Match          7.0%; Score 29.8; DB 1; Length 2702;
Best Local Similarity 51.9%; Pred. No. 3.7;
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

```

Qy  281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340
      |||  | ||||  ||  ||||  |  |||  || |||  |||  |
Db  2534 CACATCCGCACTCCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 2593

Qy  341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
      |  |  |  |||  |  |||  |  |  |||  || |||  || |||
Db  2594 AGACCTCTGTCCCTGTTCCCCTCCACTCCTCCCCTCACTCCCCTGCTCCCCCGACCACCT 2653

Qy  401 cgttctgtg 409
      |  |||  ||
Db  2654 CCTCCTCTG 2662

```

```

RESULT 8
US-08-474-379C-11
; Sequence 11, Application US/08474379C
; Patent No. 5977305
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
; TITLE OF INVENTION: PROCESSES
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,379C
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/511,715
;      FILING DATE:  20-APR-1990
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/206,188
;      FILING DATE:  01-MAR-1994
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/688,352
;      FILING DATE:  19-APR-1991
;      ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  27866/32771
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;      INFORMATION FOR SEQ ID NO:  11:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  2702 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  cDNA
;      FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  8..2701
;      FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2433
;      OTHER INFORMATION:  /note= "A shift in reading frame
;      OTHER INFORMATION:  may occur at this nucleotide."
US-08-474-379C-11

```

```

Query Match          7.0%;  Score 29.8;  DB 2;  Length 2702;
Best Local Similarity 51.9%;  Pred. No. 3.7;
Matches 67;  Conservative 0;  Mismatches 62;  Indels 0;  Gaps 0;

```

```

Qy   281  cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340
      ||||  | ||||  ||  |||||  |  |||  | || |||  | |||  |
Db   2534  CACATCCGCACTCCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 2593

Qy   341  ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
      |  |  |  | || |  | || |  |  |  | || |  || |||  |  || |||
Db   2594  AGACCTCTGTCCCTGTTCCCTCCACTCCTCCCTCACTCCCTGCTCCCCCGACCACCT 2653

Qy   401  cgttctgtg 409
      |  |  |  |
Db   2654  CCTCCTCTG 2662

```

RESULT 9  
 US-09-146-249A-11  
 ; Sequence 11, Application US/09146249A  
 ; Patent No. 6069240  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wigler, Michael H.  
 ; APPLICANT: Colicelli, John J.  
 ; TITLE OF INVENTION: Cloning by Complementation and Related  
 ; TITLE OF INVENTION: Processes  
 ; NUMBER OF SEQUENCES: 85  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/146,249A  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/511,715  
 ; FILING DATE: 20-APR-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Clough, David W.  
 ; REGISTRATION NUMBER: 36,107  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312-474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 11:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2702 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 8..2701  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 2433  
 ; OTHER INFORMATION: /note= "A shift in reading frame  
 ; OTHER INFORMATION: may occur at this nucleotide."  
 US-09-146-249A-11

Query Match 7.0%; Score 29.8; DB 3; Length 2702;  
 Best Local Similarity 51.9%; Pred. No. 3.7;

Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
Qy 281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340
    |||| | |||| | | |||| | || | || | || | || |
Db 2534 CACATCCGCACTCCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 2593

Qy 341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
    | | | | | | | | | | | | | | | | | | | | |
Db 2594 AGACCTCTGTCCCTGTTCCCCTCCACTCCTCCCCTCACTCCCCTGCTCCCCCGACCACCT 2653

Qy 401 cgttctgtg 409
    | | | |
Db 2654 CCTCCTCTG 2662
```

RESULT 10

US-08-206-188B-11

; Sequence 11, Application US/08206188B

; Patent No. 6100025

; GENERAL INFORMATION:

; APPLICANT: Wigler, Michael H.

; APPLICANT: Colicelli, John J.

; TITLE OF INVENTION: Cloning by Complementation and Related

; TITLE OF INVENTION: Processes

; NUMBER OF SEQUENCES: 84

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

; STREET: 6300 Sears Tower, 233 South Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: United States of America

; ZIP: 60606-6402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/206,188B

; FILING DATE: 01-MAR-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/511,715

; FILING DATE: 20-APR-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Clough, David W.

; REGISTRATION NUMBER: 36107

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312/474-6300

; TELEFAX: 312-474-0448

; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2702 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear



```

; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 8..2701
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2433
; OTHER INFORMATION: /note= "A shift in reading frame
; OTHER INFORMATION: may occur at this nucleotide."
US-08-206-188B-11

```

```

Query Match          7.0%; Score 29.8; DB 3; Length 2702;
Best Local Similarity 51.9%; Pred. No. 3.7;
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

```

Qy 281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340
    |||| | |||| | | |||| | || | || | || | || |
Db 2534 CACATCCGCACTCCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 2593

Qy 341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
    | | | | | | | | | | | | | | | | | | | |
Db 2594 AGACCTCTGTCCCTGTTCCCCTCCACTCCTCCCCTCACTCCCCTGCTCCCCCGACCACCT 2653

Qy 401 cgttctgtg 409
    | | | |
Db 2654 CCTCCTCTG 2662

```

```

RESULT 11
US-07-688-352C-21
; Sequence 21, Application US/07688352C
; Patent No. 5527896
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/688,352C
; FILING DATE: 19910419
; CLASSIFICATION: 435

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Borun, Michael F.
; REGISTRATION NUMBER: 25447
; REFERENCE/DOCKET NUMBER: 27805/30197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3131 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1652
; OTHER INFORMATION: /note= "A shift in reading frame
; OTHER INFORMATION: may occur at this residue."
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(743..1648, 1651..2661)
US-07-688-352C-21

```

```

Query Match          7.0%; Score 29.8; DB 1; Length 3131;
Best Local Similarity 51.9%; Pred. No. 4;
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

```

Qy   281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340
      |||| | |||| | | |||| | || | || || || || |
Db   2607 CACATCCGCACTCCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 2666

Qy   341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
      | | | | | | | | | | | | | | | | | | | | | |
Db   2667 AGACCTCTGTCCCTGTTCCCCTCCACTCCTCCCCTCACTCCCCTGCTCCCCCGACCACCT 2726

Qy   401 cgttctgtg 409
      | | | | |
Db   2727 CCTCCTCTG 2735

```

```

RESULT 12
US-09-146-249A-21
; Sequence 21, Application US/09146249A
; Patent No. 6069240
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:

```

```

;   ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET:    6300 Sears Tower, 233 South Wacker Drive
;   CITY:     Chicago
;   STATE:    Illinois
;   COUNTRY:   United States of America
;   ZIP:      60606-6402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Floppy disk
;   COMPUTER:   IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:   PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/09/146,249A
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  US 07/511,715
;   FILING DATE:   20-APR-1990
;   ATTORNEY/AGENT INFORMATION:
;   NAME:   Clough, David W.
;   REGISTRATION NUMBER:  36,107
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  312/474-6300
;   TELEFAX:   312-474-0448
;   TELEX:    25-3856
;   INFORMATION FOR SEQ ID NO: 21:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:   3131 base pairs
;   TYPE:     nucleic acid
;   STRANDEDNESS:  single
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   FEATURE:
;   NAME/KEY:  misc_feature
;   OTHER INFORMATION:  /note= "Nucleotides 429-427 and 634-670
;   OTHER INFORMATION:  may represent introns; sequence may have frame shifts
at nucleo
;   OTHER INFORMATION:  592, 1590 and 1592."
;   FEATURE:
;   NAME/KEY:  CDS
;   LOCATION:  join(2..1648, 1651..2661)
US-09-146-249A-21

```

```

Query Match          7.0%;  Score 29.8;  DB 3;  Length 3131;
Best Local Similarity 51.9%;  Pred. No. 4;
Matches 67;  Conservative 0;  Mismatches 62;  Indels 0;  Gaps 0;

```

```

Qy   281  cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340
      ||||  | ||||  ||  |||||  |  |||  | || |||  | |||  |
Db   2607  CACATCCGCACTCCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 2666

Qy   341  ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
      |  |  |  | |||  | |||  |  |  |  | |||  | |||  | |||  |
Db   2667  AGACCTCTGTCCCTGTTCCCCTCCACTCCTCCCCTCACTCCCCTGCTCCCCCGACCACCT 2726

Qy   401  cgttctgtg 409

```

| | || ||  
Db    2727 CCTCCTCTG 2735

RESULT 13  
US-08-206-188B-21  
; Sequence 21, Application US/08206188B  
; Patent No. 6100025  
; GENERAL INFORMATION:  
; APPLICANT: Wigler, Michael H.  
; APPLICANT: Colicelli, John J.  
; TITLE OF INVENTION: Cloning by Complementation and Related  
; TITLE OF INVENTION: Processes  
; NUMBER OF SEQUENCES: 84  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/206,188B  
; FILING DATE: 01-MAR-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/511,715  
; FILING DATE: 20-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clough, David W.  
; REGISTRATION NUMBER: 36107  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3131 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: /note= "Nucleotides 429-427 and 634-670 may  
; OTHER INFORMATION: represent introns; sequence may have frame shifts at  
; OTHER INFORMATION: nucleotides 328, 592, 1590 and 1592."  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: join(2..1648, 1651..2661)  
US-08-206-188B-21

Query Match 7.0%; Score 29.8; DB 3; Length 3131;  
Best Local Similarity 51.9%; Pred. No. 4;  
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
Qy 281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340
      |||| | |||| | | |||| | || | || | || | || |
Db 2607 CACATCCGCACTCCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 2666

Qy 341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
      | | | | | | | | | | | | | | | | | | | | |
Db 2667 AGACCTCTGTCCCTGTTCCCCTCCACTCCTCCCCTCACTCCCCTGCTCCCCCGACCACCT 2726

Qy 401 cgttctgtg 409
      | | | | |
Db 2727 CCTCCTCTG 2735
```

RESULT 14  
PCT-US91-02714-20  
; Sequence 20, Application PC/TUS9102714  
; GENERAL INFORMATION:  
; APPLICANT: Wigler, Michael H.  
; APPLICANT: Colicelli, John J.  
; TITLE OF INVENTION: Cloning by Complementation and Related  
; TITLE OF INVENTION: Processes  
; NUMBER OF SEQUENCES: 55  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
; ADDRESSEE: Bicknell  
; STREET: Two First National Plaza, 20 South Clark  
; STREET: Street  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US91/02714  
; FILING DATE: 19910419  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/511,715  
; FILING DATE: 20-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Borun, Michael F.  
; REGISTRATION NUMBER: 25447  
; REFERENCE/DOCKET NUMBER: 27805/30197  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 346-5750  
; TELEFAX: (312) 984-9740  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 20:

```

; SEQUENCE CHARACTERISTICS:
;   LENGTH: 3131 base pairs
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 743..1651
; FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1652
;   OTHER INFORMATION: /note= "A shift in reading frame
;   OTHER INFORMATION: may occur at this residue."
PCT-US91-02714-20

```

```

Query Match          7.0%; Score 29.8; DB 5; Length 3131;
Best Local Similarity 51.9%; Pred. No. 4;
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

```

Qy  281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340
      ||||  | ||||  ||  ||||  |  |||  || ||||  |||  |
Db  2607 CACATCCGCACTCCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 2666

Qy  341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
      |  |  |  |||  |  |||  |  |  |||  || ||||  ||  ||||
Db  2667 AGACCTCTGTCCCTGTTCCCCTCCACTCCTCCCCTCACTCCCCTGCTCCCCCGACCACCT 2726

Qy  401 cgttctgtg 409
      |  |  ||  ||
Db  2727 CCTCCTCTG 2735

```

```

RESULT 15
US-08-474-379C-64
; Sequence 64, Application US/08474379C
; Patent No. 5977305
; GENERAL INFORMATION:
;   APPLICANT: Wigler, Michael H.
;   APPLICANT: Colicelli, John J.
;   TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
;   TITLE OF INVENTION: PROCESSES
;   NUMBER OF SEQUENCES: 88
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;     STREET: 233 South Wacker Drive/6300 Sears Tower
;     CITY: Chicago
;     STATE: Illinois
;     COUNTRY: United States of America
;     ZIP: 60606-6402
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/474,379C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/206,188
; FILING DATE: 01-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/688,352
; FILING DATE: 19-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 27866/32771
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3705 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 116..2773
US-08-474-379C-64

```

```

Query Match          7.0%; Score 29.8; DB 2; Length 3705;
Best Local Similarity 51.9%; Pred. No. 4.3;
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

```

Qy  281 cacactgccgctccccacgctaaatttgggggctacagtgcacacgctagccgattaac 340
      ||||  | ||||  ||  ||||  |  ||  | || ||  ||  |
Db  2722 CACATCCGCACTCCCAGCTCCTGGTGGCGGGGGGTCAGGTGGAGACCCTACCTGATCCCC 2781

Qy  341 ggctcacgctaccaggcgctctacgcggatgtgccccctagccagcttctctctccccct 400
      |  |  |  | ||  |  ||  |  |  |  |  |  |  |  |  |  |  |  |
Db  2782 AGACCTCTGTCCCTGTTCCCCTCCACTCCTCCCCTCACTCCCCTGCTCCCCCGACCACCT 2841

Qy  401 cgttctgtg 409
      |  |  |  |
Db  2842 CCTCCTCTG 2850

```

```

Search completed: February 7, 2002, 10:51:52
Job time: 6078 sec

```

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 08:20:45 ; Search time 4942.22 Seconds  
(without alignments)  
924.070 Million cell updates/sec

Title: US-09-394-745-6332  
Perfect score: 425  
Sequence: 1 cggacgcgtgggtgcaattt.....tgtgggtgcctctctcaacct 425

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_htc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_htc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	No.	Score	Query Match	% Length	ID	Description
c	1	409.8	96.4	738	11 BG837751	BG837751 Zm10_05b0



c	2	205.6	48.4	330	11	BG349846	BG349846	947031G10
c	3	205.6	48.4	332	11	BG349673	BG349673	947031G10
c	4	205.6	48.4	337	11	BG349674	BG349674	947031G10
	5	205.2	48.3	427	11	BG550269	BG550269	947039B02
	6	200.8	47.2	473	10	AW287785	AW287785	829008D03
c	7	200	47.1	296	11	BG360891	BG360891	947043D12
c	8	197.2	46.4	372	11	BG840972	BG840972	MEST14-B0
c	9	197.2	46.4	410	11	BG842137	BG842137	MEST36-F0
c	10	197	46.4	379	11	BG874098	BG874098	MEST46-C0
	11	187.2	44.0	371	11	BG840656	BG840656	MEST14-B0
	12	185.8	43.7	208	11	BG349675	BG349675	947031G10
	13	182.2	42.9	351	11	BG349676	BG349676	947031G10
	14	180	42.4	216	10	AA072465	AA072465	zEST00696
	15	142.4	33.5	256	11	BG355157	BG355157	947043D12
c	16	131.2	30.9	290	11	BG549642	BG549642	947080B10
	17	128.4	30.2	343	11	BG549643	BG549643	947080B10
c	18	128	30.1	279	11	BG549998	BG549998	947080B10
	19	82.8	19.5	289	11	BG549999	BG549999	947080B10
c	20	35.2	8.3	622	13	AZ857609	AZ857609	2M0162A06
	21	34.8	8.2	343	11	BF199292	BF199292	249419 MA
c	22	34.8	8.2	960	11	BG759439	BG759439	602711981
	23	34.2	8.0	228	11	BF511953	BF511953	UI-H-BI4-
	24	34.2	8.0	335	11	BF511954	BF511954	UI-H-BI4-
	25	34.2	8.0	394	10	AW974268	AW974268	EST386371
	26	34.2	8.0	459	10	AI459082	AI459082	tk03g12.x
c	27	34.2	8.0	523	10	AI737962	AI737962	606044A09
c	28	33.6	7.9	544	13	AQ578052	AQ578052	nbxb00920
c	29	33.6	7.9	555	13	AQ289739	AQ289739	nbxb0035J
c	30	33.6	7.9	559	13	AQ689991	AQ689991	nbxb0081E
c	31	33.6	7.9	594	13	AQ290298	AQ290298	nbxb0036N
c	32	33.6	7.9	631	13	AQ843663	AQ843663	nbxb0024L
c	33	33.6	7.9	701	13	AQ578229	AQ578229	nbxb0092D
c	34	33.6	7.9	779	13	AQ689737	AQ689737	nbxb0080O
c	35	33.6	7.9	783	13	AQ365723	AQ365723	nbxb0064I
	36	33.4	7.9	204	10	BE668096	BE668096	156519 MA
	37	33.2	7.8	450	10	AU198120	AU198120	AU198120
	38	32.8	7.7	1056	13	CNS03R34	AL256729	Tetraodon
	39	32.6	7.7	258	10	AW445516	AW445516	81732 MAR
	40	32.6	7.7	456	10	AU198080	AU198080	AU198080
	41	32.6	7.7	491	11	BF655640	BF655640	FM1_46_H1
	42	32.4	7.6	345	10	AU090517	AU090517	AU090517
	43	32.4	7.6	412	10	AA647877	AA647877	vn08b10.r
	44	32.4	7.6	576	10	AA154381	AA154381	mr32c03.r
	45	32.4	7.6	1101	13	CNS01720	AL107514	Drosophil

# ALIGNMENTS

RESULT 1  
BG837751/c  
LOCUS BG837751 738 bp mRNA EST 25-MAY-2001  
DEFINITION Zm10\_05b06\_A Zm10\_AAFC\_ECORC\_Fusarium\_graminearum\_corn\_silk Zea  
mays cDNA clone Zm10\_05b06, mRNA sequence.  
ACCESSION BG837751  
VERSION BG837751.1 GI:14204074  
KEYWORDS EST.

SOURCE Zea mays.  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 738)  
 AUTHORS Harris,L.J., Balcerzak,M., Allard,S., Saparno,A., Couroux,P., De  
 Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S., Singh,J.A, Sprott  
 ,D. and Tinker,N.A.  
 TITLE Expressed Sequence Tags from Maize Silk Six Hours After Silk  
 Channel Inoculation with Fusarium graminearum  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Harris, Linda J.  
 Eastern Cereal and Oilseed Research Centre  
 Agriculture and Agri-food Canada  
 Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,  
 CANADA  
 Tel: (613) 759-1314  
 Fax: (613) 759-6566  
 Email: harrislj@em.agr.ca.  
 FEATURES Location/Qualifiers  
 source 1. .738  
 /organism="Zea mays"  
 /cultivar="CO388"  
 /db\_xref="taxon:4577"  
 /clone="Zm10\_05b06"  
 /clone\_lib="Zm10\_AAFC\_ECORC\_Fusarium\_graminearum\_corn\_silk  
 "  
 /tissue\_type="Silk"  
 /dev\_stage="4-5 days post-silk emergence"  
 /note="Vector: Bluescript SK+/XhoI-EcoRI; Site\_1: EcoRI;  
 Site\_2: XhoI; Field-grown corn was silk channel-inoculated  
 in the morning (~10 am) with 1 ml of a macroconidial  
 suspension (500,000 spores/ml) of Fusarium graminearum and  
 silk channels were collected and immediately frozen in  
 liquid nitrogen 6 hours later. RNA was extracted from  
 silk tissue between 1 cm below and above the inoculation  
 point in the silk channel, RNA from five silk channels was  
 pooled."  
 BASE COUNT 187 a 207 c 161 g 183 t  
 ORIGIN

Query Match 96.4%; Score 409.8; DB 11; Length 738;  
 Best Local Similarity 99.5%; Pred. No. 1.2e-112;  
 Matches 411; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy      13  tgcaatttgaggagagagacgagatcatgaggaagcaatactcccctgtgctctacttct 72
          |||
Db      725  TGCAATTTGAGGAGAGAGACGAGATCATGAGGAAGCAATACTCCCCTGTGCTCTACTTCT 666

Qy      73  gcctgatggcccttgctcgtagctgctatggtctgtgtcatgtacaccacctcggcacaag 132
          |||
Db      665  GCCTGATGGCCCTTGTCGTAGCTGCTATGGTCTGTGTCATGTACACCACCTCGGCACAAG 606

Qy     133  caggaaggagtggctacaactcgtacgaacctgatggaaggggtggatacaactctgttc 192
          |||

```

```

Db      605 CAGGAAGGAGTGGCTACAACCTCGTACGAACCTGATGGAAGGGGTGGATACAACCTCTGTTC 546
Qy      193 ccatcaacggcggtggcagcccctagctaggcggtggatccgagcctgtatcagaaatcg 252
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      545 CCATCAACGGCGGTGGCAGCCCCTAGCTAGGCGGTGGATCCGAGCCTGTATCAGAAATCG 486
Qy      253 aaataatataagactgtcttcaacggatcacactgccgctccccacgctaaatttgggg 312
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      485 AAATAATATAAGACTGTCTTCAACGGATCACACTGCCGCTCCCCACGCTAAATTTGGGG 426
Qy      313 gctacagtgcacacgctagccgattaacggctcacgctaccaggcgctctacgcggtatgt 372
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      425 GCTACAGTGCACACGCTAGCCGATCAACGGCTCACGCTACCAGGCGCTCTACGCGGATGT 366
Qy      373 gccccctagccagcttctctctctccccctcgttctgtggtgcctctctcaacct 425
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      365 GCCCCCTAGCCAGCTTCTCTCTCCCCCTCGTTCTGTGGTGCCTCTCTACCCT 313

```

```

RESULT      2
BG349846/c
LOCUS       BG349846      330 bp      mRNA                      EST           01-MAR-2001
DEFINITION  947031G10.x3 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA
            sequence.
ACCESSION   BG349846
VERSION     BG349846.1  GI:13178588
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 330)
AUTHORS    Walbot,V.
TITLE      Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL     Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 947031 row: G column: 10.
FEATURES             Location/Qualifiers
     source            1. .330
                       /organism="Zea mays"
                       /cultivar="B73"
                       /db_xref="taxon:4577"
                       /clone_lib="947 - 2 week shoot from Barkan lab"
                       /tissue_type="leaf and stem, including leaf base"
                       /dev_stage="2 week old seedling (3 leaves)"
                       /lab_host="XL1-Blue"
                       /note="Organ: shoot; Vector: Lambda ZAP (pBlueScript SK-);
                       Site_1: EcoRI; Site_2: XhoI; Directionally cloned using
                       Stratagene's UniZap XR cDNA cloning kit with the 5' end

```

BASE COUNT	83 a	93 c	69 g	85 t
ORIGIN				

[illegible]

```

RESULT      3
BG349673/c
LOCUS       BG349673          332 bp      mRNA                      EST              01-MAR-2001
DEFINITION  947031G10.x1 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA
sequence.
ACCESSION   BG349673
VERSION     BG349673.1   GI:13178400
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 332)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 947031 row: G column: 10.
FEATURES             Location/Qualifiers
   source             1. .332
                     /organism="Zea mays"

```



Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 947031 row: G column: 10.

FEATURES  
source Location/Qualifiers  
1. .337  
/organism="Zea mays"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone\_lib="947 - 2 week shoot from Barkan lab"  
/tissue\_type="leaf and stem, including leaf base"  
/dev\_stage="2 week old seedling (3 leaves)"  
/lab\_host="XL1-Blue"  
/note="Organ: shoot; Vector: Lambda ZAP (pBlueScript SK-);  
Site\_1: EcoRI; Site\_2: XhoI; Directionally cloned using  
Stratagene's UniZap XR cDNA cloning kit with the 5' end  
at the EcoRI site. The library represents 8 x 10e5  
independent recombinant phage. The plants were greenhouse  
grown."  
BASE COUNT 84 a 95 c 70 g 88 t  
ORIGIN

Query Match 48.4%; Score 205.6; DB 11; Length 337;  
Best Local Similarity 94.2%; Pred. No. 2.6e-51;  
Matches 226; Conservative 0; Mismatches 9; Indels 5; Gaps 1;

```
Qy      24 gagagagacgagatcatgaggaagcaataactcccctgtgctctacttctgcctgatggcc 83
      | |||||
Db      331 GGGAGAGACGAGATCATGAGGAAGCAATACTCCCCTGTGCTCTCCTTGTGCCTGATGGCC 272

Qy      84 cttgtcgtagctgctatggtctgtgtcatgtacaccacctcggcacaagcaggaaggagt 143
      | |||||
Db      271 CTTGTCTAGCTGCTATGGTCTGTGTCATGTACACCACCTCGGCACAAGCAGGAAGGAGT 212

Qy     144 ggctacaactcgtagcgaacctgatggaagggtggatacaactctgttcccatcaacggc 203
      | |||||
Db     211 GGCTACAACCTCGTACGAACCTGATGGAAGGAGTGGATACAACTCTGTTCCAATAAACGGC 152

Qy     204 ggtggcagcccctagctaggcggtggatccgagcctgtatcagaaatcgaaataatataa 263
      | |||||
Db     151 GGTGGCAGCCCCTAGCCAGGCGGTG-----GAGCCTGTATCAGAAATCAAAAAAATATAA 97
```

RESULT 5  
BG550269  
LOCUS BG550269 427 bp mRNA EST 05-APR-2001  
DEFINITION 947039B02.y1 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA  
sequence.  
ACCESSION BG550269  
VERSION BG550269.1 GI:13558914  
KEYWORDS EST.  
SOURCE Zea mays.  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 427)

AUTHORS Walbot,V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford  
 University

JOURNAL Unpublished (1999)

COMMENT Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Plate: 947039 row: B column: 02.

FEATURES Location/Qualifiers

source 1. .427  
 /organism="Zea mays"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone\_lib="947 - 2 week shoot from Barkan lab"  
 /tissue\_type="leaf and stem, including leaf base"  
 /dev\_stage="2 week old seedling (3 leaves)"  
 /lab\_host="XL1-Blue"  
 /note="Organ: shoot; Vector: Lambda ZAP (pBlueScript SK-);  
 Site\_1: EcoRI; Site\_2: XhoI; Directionally cloned using  
 Stratagene's UniZap XR cDNA cloning kit with the 5' end  
 at the EcoRI site. The library represents 8 x 10e5  
 independent recombinant phage. The plants were greenhouse  
 grown."

BASE COUNT 126 a 80 c 104 g 117 t

ORIGIN

Query Match 48.3%; Score 205.2; DB 11; Length 427;  
 Best Local Similarity 94.5%; Pred. No. 3.8e-51;  
 Matches 225; Conservative 0; Mismatches 8; Indels 5; Gaps 1;

Qy 26 gagagacgagatcatgaggaagcaataactcccctgtgctctacttctgcctgatggccct 85  
 |||

Db 2 GAGAGACGAGATCATGAGGAAGCAATACTCCCCTGTGCTCTCCTTGTGCCTGATGGCCCT 61

Qy 86 tgtcgtagctgctatgggtctgtgtcatgtacaccacctcggcacaagcaggaaggagtgg 145  
 |||

Db 62 TGTCGTAGCTGCTATGGTCTGTGTCATGTACACCACCTCGGCACAAGCAGGAAGGAGTGG 121

Qy 146 ctacaactcgtaacgaacctgatggaaggggtggatacaactctgttcccatcaacggcgg 205  
 |||

Db 122 CTACAACTCGTACGAACCTGATGGAAGGAGTGGATACAACTCTGTTCCAATAAACGGCGG 181

Qy 206 tggcagcccctagctaggcgggtggatccgagcctgtatcagaaatcgaaataatataa 263  
 |||

Db 182 TGGCAGCCCCTAGCCAGGCGGTG-----GAGCCTGTATCAGAAATCAAAAAAATATAA 234

RESULT 6  
 AW287785

LOCUS AW287785 473 bp mRNA EST 09-FEB-2000  
 DEFINITION 829008D03.x1 829 - Silk infected with Fusarium Zea mays cDNA, mRNA sequence.  
 ACCESSION AW287785  
 VERSION AW287785.1 GI:6681798  
 KEYWORDS EST.  
 SOURCE Zea mays.  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 473)  
 AUTHORS Walbot,V.  
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Plate: 829008 row: D column: 03.  
 FEATURES Location/Qualifiers  
 source 1. .473  
 /organism="Zea mays"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone\_lib="829 - Silk infected with Fusarium"  
 /tissue\_type="silk"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: silk; Vector: pBluescript II XR; Site\_1: XhoI; Site\_2: EcoRI; cDNA library of silks infected with 1 microliter of 500,000 spores/ml solution of Fusarium graminearum DAOM 180378. Prepared by Sharon Allard of Eastern Cereal and Oilseed Research Centre, Agriculture and Agri-Food Canada using Stratagene cDNA synthesis kit. Silk was harvested at 72 hours p.i."  
 BASE COUNT 148 a 81 c 143 g 101 t  
 ORIGIN

Query Match 47.2%; Score 200.8; DB 10; Length 473;  
 Best Local Similarity 92.9%; Pred. No. 8.3e-50;  
 Matches 223; Conservative 0; Mismatches 12; Indels 5; Gaps 1;

Qy 24 gagagagacgagatcatgaggaagcaataactcccctgtgctctacttctgcctgatggcc 83  
 ||||| |||||  
 Db 81 GAGAGAGAGAGAATCATGAGGAAGCAATACTCCCCTGTGCTCTCCTTGTGCCTGATGGCC 140  
 Qy 84 cttgtcgtagctgctatggtctgtgtcatgtacaccacctcggcacaagcaggaaggagt 143  
 ||||| |||||  
 Db 141 CTTGTCTAGCTGCTATGGTCTGTGTCATGTACACCACCTCGGCACAAGCAGGAAGGAGT 200  
 Qy 144 ggctacaactcgtacgaacctgatggaaggggtggatacaactctgttcccatcaacggc 203



```

      |||
Db    201 GGCTACAACCTCGTACGAACCTGATGGAAGGAGTGGATACAACCTCTGTTCCAATAAACGGC 260

Qy    204 ggtggcagcccctagctaggcggtggatccgagcctgtatcagaaatcgaaataatataa 263
      |||
Db    261 GGTGGCAGCCCCTAGCCAGGCGGTG-----GAGCCTGTATCAGAAATCAAAAAAATATAA 315

```

RESULT 7

BG360891/c

LOCUS BG360891 296 bp mRNA EST 08-MAR-2001  
 DEFINITION 947043D12.x2 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA  
 sequence.

ACCESSION BG360891

VERSION BG360891.1 GI:13249988

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 296)

AUTHORS Walbot,V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford  
 University

JOURNAL Unpublished (1999)

COMMENT Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Plate: 947043 row: D column: 12.

FEATURES

source Location/Qualifiers  
 1..296  
 /organism="Zea mays"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone\_lib="947 - 2 week shoot from Barkan lab"  
 /tissue\_type="leaf and stem, including leaf base"  
 /dev\_stage="2 week old seedling (3 leaves)"  
 /lab\_host="XL1-Blue"  
 /note="Organ: shoot; Vector: Lambda ZAP (pBlueScript SK-);  
 Site\_1: EcoRI; Site\_2: XhoI; Directionally cloned using  
 Stratagene's UniZap XR cDNA cloning kit with the 5' end  
 at the EcoRI site. The library represents 8 x 10e5  
 independent recombinant phage. The plants were greenhouse  
 grown."

BASE COUNT 72 a 81 c 66 g 77 t

ORIGIN

Query Match 47.1%; Score 200; DB 11; Length 296;  
 Best Local Similarity 93.6%; Pred. No. 1.2e-49;  
 Matches 221; Conservative 0; Mismatches 10; Indels 5; Gaps 1;

Qy	28	gagacgagatcatgaggaagcaataactcccctgtgctctacttctgcctgatggcccttg	87
Db	292	GAGACGAGATCATGAGGAAGCAATACTCCCCTGTGCTCTCCTTGTGCCTGATGGCCCTTG	233
Qy	88	tcgtagctgctatgggtctgtgtcatgtacaccacctcggcacaagcaggaaggagtggct	147
Db	232	TCGTAGCTGCTATGGTCTGTGTCATGTACACCACCTCGGCACAAGCAGGAAGGAGTGGCT	173
Qy	148	acaactcgtacgaacctgatggaaggggtggatacaactctgttcccatcaacggcgggtg	207
Db	172	ACAACTCGTACGAACCTGATGGAAGGAGTGGATACAACTCTGTTACAATAACCGGCGGTG	113
Qy	208	gcagcccctagctaggcgggtggatccgagcctgtatcagaaatcgaaataatataa	263
Db	112	GCAGCCCCTAGCCAGGCGGTG-----GAGCCTGTATCAGAAATCAAAAAAATATAA	62



G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA

Tel: 515-294-0975

Fax: 515-294-2299

Email: schnable@iastate.edu

PCR Primers

FORWARD: T7-1 (AA TAC GAC TCA CTA TAG)

BACKWARD: T3 (ATT AAC CCT CAC TAA AG)

Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

FEATURES  
source Location/Qualifiers  
1. .410  
/organism="Zea mays"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="MEST36-F07"  
/clone\_lib="ISUM3-TL"  
/tissue\_type="Seedling and silk"  
/lab\_host="DH10B"  
/note="Vector: pT7T3PAC; Site\_1: EcoRI; Site\_2: NotI;  
ds-cDNA molecules were generated as follows. First-strand  
cDNA was prepared from oligo-dT selected mRNA by priming  
with a NotI oligo-dT primer (5'  
AACTGGAAGAATTCGCGGCCGAGGAATTTTTTTTTTTTTTTTTT). The  
resulting DNA:RNA hybrid was treated with RNase H and used  
as a template for DNA PolI-catalyzed second strand  
synthesis. After the addition of EcoRI adaptors, the  
ds-cDNAs were digested with NotI and size-selected. The  
resulting molecules were directionally cloned into the  
EcoRI and NotI sites of the pT7T3PAC vector."  
BASE COUNT 114 a 99 c 79 g 118 t  
ORIGIN

Query Match 46.4%; Score 197.2; DB 11; Length 410;  
Best Local Similarity 94.3%; Pred. No. 9.5e-49;  
Matches 217; Conservative 0; Mismatches 8; Indels 5; Gaps 1;

Qy 34 agatcatgaggaagcaataactcccctgtgctctacttctgcctgatggcccttgctcgtag 93  
|||||  
Db 410 AGATCATGAGGAAGCAATACTCCCCTGTGCTCTCCTTGTGCCTGATGGCCCTTGTCGTAG 351  
  
Qy 94 ctgctatggtctgtgtcatgtacaccacctcggcacaagcaggaaggagtggctacaact 153  
|||||  
Db 350 CTGCTATGGTCTGTGTCATGTACACCACCTCGGCACAAGCAGGAAGGAGTGGCTACAAC 291  
  
Qy 154 cgtacgaacctgatggaaggggtggatatacaactctgttcccatcaacggcggtggcagcc 213  
|||||  
Db 290 CGTACGAACCTGATGGAAGGAGTGGATACAACCTCTGTTCCAATAAACGGCGGTGGCAGCC 231  
  
Qy 214 cctagctaggcgggtggatccgagcctgtatcagaaatcgaaataatataa 263  
|||||  
Db 230 CCTAGCCAGGCGGTG-----GAGCCTGTATCAGAAATCAAAAAATATAA 186

RESULT 10  
BG874098/c  
LOCUS BG874098 379 bp mRNA EST 29-MAY-2001  
DEFINITION MEST46-C08.T3 ISUM4-TN Zea mays cDNA clone MEST46-C08 3', mRNA

sequence.

ACCESSION BG874098

VERSION BG874098.1 GI:14245516

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 379)

AUTHORS Qiu, F., Cui, F., Guo, L., Ashlock, D.A., Wen, T.J. and Schnable, P.S.

TITLE Expressed Sequence Tags from B73 Maize Seedlings and Silks

JOURNAL Unpublished (2001)

COMMENT Contact: Patrick S. Schnable  
Schnable Laboratory  
Iowa State University  
G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA  
Tel: 515-294-0975  
Fax: 515-294-2299  
Email: schnable@iastate.edu  
PCR Primers  
FORWARD: T7-1 (AA TAC GAC TCA CTA TAG)  
BACKWARD: T3 (ATT AAC CCT CAC TAA AG)  
Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

FEATURES Location/Qualifiers

source 1. .379  
/organism="Zea mays"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="MEST46-C08"  
/clone\_lib="ISUM4-TN"  
/tissue\_type="Seedling and silk"  
/lab\_host="DH10B"  
/note="Vector: pT7T3PAC; Site\_1: EcoRI; Site\_2: NotI;  
ds-cDNA molecules were generated as follows. First-strand  
cDNA was prepared from oligo-dT selected mRNA by priming  
with a NotI oligo-dT primer (5'  
AACTGGAAGAATTGCGGCCGCGAGGAATTTTTTTTTTTTTTTTTT). The  
resulting DNA:RNA hybrid was treated with RNase H and used  
as a template for DNA PolI-catalyzed second strand  
synthesis. After the addition of EcoRI adaptors, the  
ds-cDNAs were digested with NotI and size-selected. The  
resulting molecules were directionally cloned into the  
EcoRI and NotI sites of the pT7T3PAC vector. The library  
then went through one round of normalization to CoT value  
of 5 based on the methods of Marcelo Bento Soares (Genome  
Research 6: 791-806, 1996)."

BASE COUNT 90 a 94 c 74 g 121 t

ORIGIN

Query Match 46.4%; Score 197; DB 11; Length 379;  
Best Local Similarity 93.6%; Pred. No. 1.1e-48;  
Matches 218; Conservative 0; Mismatches 10; Indels 5; Gaps 1;

Qy 31 acgagatcatgaggaagcaatactcccctgtgctctacttctgcctgatggcccttgctcg 90  
|||||

```

Db      379 ACGAGATCATGAGGAAGCAATACTCCCCTGTGCTCTCCTTGTGCCTGATGGCCCTTGTCG 320
Qy      91 tagctgctatggtctgtgtcatgtacaccacctcggcacaagcaggaaggagtggctaca 150
        |||
Db      319 TAGCTGCTATGGTCTGTGTCATGTACACCACCTCGGCACAAGCAGGAAGGAGTGGCTACA 260
Qy      151 actcgtacgaacctgatggaaggggtggatacaactctgttcccatcaacggcggtggca 210
        |||
Db      259 ACTCGTACGAACCTGATGGAAGGAGTGGATACAACCTTTGTTCCAATAAACGGCGGTGGCA 200
Qy      211 gcccctagctaggcggtggatccgagcctgtatcagaaatcgaaataatataa 263
        |||
Db      199 GCCCCAAGCCAGGCGGTG-----GAGCCTGTATCAGAAATCAAAAAATATAA 152

```

# RESULT 11

BG840656

LOCUS BG840656 371 bp mRNA EST 29-MAY-2001

DEFINITION MEST14-B02.T7-1 ISUM4-TN Zea mays cDNA clone MEST14-B02 5', mRNA sequence.

ACCESSION BG840656

VERSION BG840656.2 GI:14242839

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 371)

AUTHORS Qiu, F., Cui, F., Guo, L., Ashlock, D.A., Wen, T.J. and Schnable, P.S.

TITLE Expressed Sequence Tags from B73 Maize Seedlings and Silks

JOURNAL Unpublished (2001)

COMMENT On May 25, 2001 this sequence version replaced gi:14206978.

Contact: Patrick S. Schnable

Schnable Laboratory

Iowa State University

G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA

Tel: 515-294-0975

Fax: 515-294-2299

Email: schnable@iastate.edu

PCR Primers

FORWARD: T7-1 (AA TAC GAC TCA CTA TAG)

BACKWARD: T3 (ATT AAC CCT CAC TAA AG)

Seq primer: primer T7-1 (AA TAC GAC TCA CTA TAG).

## FEATURES

source

Location/Qualifiers

1. .371

/organism="Zea mays"

/cultivar="B73"

/db\_xref="taxon:4577"

/clone="MEST14-B02"

/clone\_lib="ISUM4-TN"

/tissue\_type="Seedling and silk"

/lab\_host="DH10B"

/note="Vector: pT7T3PAC; Site\_1: EcoRI; Site\_2: NotI;

ds-cDNA molecules were generated as follows. First-strand

cDNA was prepared from oligo-dT selected mRNA by priming

with a NotI oligo-dT primer (5'

BASE COUNT	110 a	75 c	93 g	93 t
ORIGIN	.			

Qy	44	gaagcaataactccccctgtgctctacttctgcoctgatggcccttgtcgtagctgctatgggt	103
Db	10	GAAGCAATACTCCCCTGTGCTCTCCTTGTGCCTGATGGCCCTTGTTCGTAGCTGCTATGGT	69
Qy	104	ctgtgtcatgtacaccacctcggcacaagcaggaaggagtggctacaactcgtacgaacc	163
Db	70	CTGTGTCATGTACACCACCTCGGCACAAGCAGGAAGGAGTGGCTACAACCTCGTACGAACC	129
Qy	164	tgatggaagggtggatacaactctgttcccatcaacggcggtggcagcccctagctagg	223
Db	130	TGATGGAAGGAGTGGATACAACTCTGTTCCAATAAACGGCGGTGGCAGCCCCTAGCCAGG	189
Qy	224	cgggtggatccgagcctgtatcagaaatcgaaataatataa	263
Db	190	CGGTG-----GAGCCTGTATCAGAAATCAAAAAAATATAA	224

```

RESULT 12
BG349675
LOCUS BG349675 208 bp mRNA EST 01-MAR-2001
DEFINITION 947031G10.y1 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA
sequence.
ACCESSION BG349675
VERSION BG349675.1 GI:13178402
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 208)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA

```

```

FEATURES             Location/Qualifiers
     source            1. .208
                        /organism="Zea mays"
                        /cultivar="B73"
                        /db_xref="taxon:4577"
                        /clone_lib="947 - 2 week shoot from Barkan lab"
                        /tissue_type="leaf and stem, including leaf base"
                        /dev_stage="2 week old seedling (3 leaves)"
                        /lab_host="XL1-Blue"
                        /note="Organ: shoot; Vector: Lambda ZAP (pBlueScript SK-);
                        Site_1: EcoRI; Site_2: XhoI; Directionally cloned using
                        Stratagene's UniZap XR cDNA cloning kit with the 5' end
                        at the EcoRI site. The library represents 8 x 10e5
                        independent recombinant phage. The plants were greenhouse
                        grown."
BASE COUNT            49 a           55 c           58 g           46 t
ORIGIN

```

Query Match 43.7%; Score 185.8; DB 11; Length 208;  
Best Local Similarity 96.4%; Pred. No. 2e-45;  
Matches 190; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy	37	tc	atgaggaagcaataactcccctgtgctctacttctgcctgatggcccttgctcgtagctg	96
Db	1	TC	ATGAGGAAGCAATACTCCCCTGTGCTCTCCTTGTGCCTGATGGCCCTTGTCGTAGCTG	60
Qy	97	ct	atgggtctgtgtcatgtacaccacctcggcacaagcaggaaggagtggctacaactcgt	156
Db	61	CT	ATGGTCTGTGTCATGTACACCACCTCGGCACAAGCAGGAAGGAGTGGCTACAACCTCGT	120
Qy	157	ac	gaacctgatggaaggggtggatacaactctgttcccatcaacggcggtggcagcccct	216
Db	121	AC	GAACCTGATGGAAGGAGTGGATACAACCTCTGTTCCAATAAACGGCGGTGGCAGCCCCCT	180
Qy	217	ag	ctaggcggtggatcc	233
Db	181	AG	CCAGGCGGTGGAGCC	197

```

RESULT 13
BG349676
LOCUS BG349676 351 bp mRNA EST 01-MAR-2001
DEFINITION 947031G10.y2 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA
sequence.
ACCESSION BG349676
VERSION BG349676.1 GI:13178403
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.

```



REFERENCE 1 (bases 1 to 351)  
AUTHORS Walbot,V.  
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University  
JOURNAL Unpublished (1999)  
COMMENT Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 947031 row: G column: 10.

FEATURES  
source Location/Qualifiers  
1. .351  
/organism="Zea mays"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone\_lib="947 - 2 week shoot from Barkan lab"  
/tissue\_type="leaf and stem, including leaf base"  
/dev\_stage="2 week old seedling (3 leaves)"  
/lab\_host="XL1-Blue"  
/note="Organ: shoot; Vector: Lambda ZAP (pBlueScript SK-); Site\_1: EcoRI; Site\_2: XhoI; Directionally cloned using Stratagene's UniZap XR cDNA cloning kit with the 5' end at the EcoRI site. The library represents 8 x 10e5 independent recombinant phage. The plants were greenhouse grown."

BASE COUNT 97 a 73 c 92 g 89 t  
ORIGIN

Query Match 42.9%; Score 182.2; DB 11; Length 351;  
Best Local Similarity 93.8%; Pred. No. 3e-44;  
Matches 213; Conservative 0; Mismatches 8; Indels 6; Gaps 2;

Qy 38 catgaggaagcaataactcccctgtgctctacttctgcctgatggcccttgctcgtagctgc 97  
|||||  
Db 1 CATGAGGAAGCAATACTCCCCTGTGCTCTCCTTGTGCCTGATGGCCCTTGTCGTAGCTGC 60

Qy 98 tatggtctgtgtcatgtacaccacctcggcacaagca-ggaaggagtggctacaactcgt 156  
|||||  
Db 61 TATGGTCTGTGTATGTACACCACTCGGCACAAGCACGGAAGGAGTGGCTACAACCTCGT 120

Qy 157 acgaacctgatggaaggggtggatacaactctgttcccatcaacggcggtggcagccoct 216  
|||||  
Db 121 ACGAACCTGATGGAAGGAGTGGATACAACTCTGTTCCAATAAACGGCGGTGGCAGCCCCT 180

Qy 217 agctaggcggtggatccgagcctgtatcagaaatcgaaataatataa 263  
|||  
Db 181 AGCCAGGCGGTG-----GAGCCTGTATCAGAAATCAAAAAATATAA 222

RESULT 14  
AA072465  
LOCUS AA072465 216 bp mRNA EST 02-OCT-1996  
DEFINITION zEST00696 Maize Leaf, Stratagene #937005 Zea mays cDNA clone

csuh00696 5' end, mRNA sequence.

ACCESSION AA072465

VERSION AA072465.1 GI:1590803

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 216)

AUTHORS Baysdorfer,C.

TITLE The Maize cDNA Program

JOURNAL Unpublished (1993)

COMMENT Contact: Baysdorfer C  
California State University  
Dept Biol Sci, California State Univ, Hayward, CA 94542  
Tel: 5108853459  
Fax: 5108854747  
Email: cbaysdor@haywire.csuhayward.edu  
Seq primer: SK.

FEATURES Location/Qualifiers

source 1. .216  
/organism="Zea mays"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="csuh00696"  
/clone\_lib="Maize Leaf, Stratagene #937005"  
/note="Vector: Uni-ZAP; Site\_1: EcoR1; Site\_2: Xho1; mRNA  
isolated from illuminated leaves and sheaths of 5 week old  
plant. cDNA directionally cloned into vector. "

BASE COUNT 57 a 54 c 56 g 47 t 2 others

ORIGIN

Query Match 42.4%; Score 180; DB 10; Length 216;  
Best Local Similarity 89.7%; Pred. No. 1.1e-43;  
Matches 192; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 36 atcatgaggaagcaataactcccctgtgctctacttctgcctgatggcccttgctgtagct 95  
|||||

Db 1 ATCATGAGGAAGCAATACTCCCCTGTGCTCTCCTTGTGCCTGATGNNCCTTGTCGTAGCT 60

Qy 96 gctatggtctgtgtcatgtacaccacctcggcacaagcaggaaggagtggctacaactcg 155  
|||||

Db 61 GCTATGGTCTGTGTCATGTACACCACCTCGGCACAAGCAGGAAGGAGTGGCTACAACCTCG 120

Qy 156 tacgaacctgatggaaggggtggatacaactctgttcccatcaacggcgggtggcagcccc 215  
|||||

Db 121 TACGAACCTGATGGAAGGAGTGGATACAACTCTGTTCCAATAAACGGCGGTGGCAGCCCC 180

Qy 216 tagctaggcgggtggatccgagcctgtatcagaaa 249  
|||||

Db 181 TAGCCAGGCGTGGAGCTGTATCAGAAATCAAAAA 214

RESULT 15  
BG355157

```

LOCUS       BG355157             256 bp      mRNA      EST      06-MAR-2001
DEFINITION  947043D12.y1 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA
sequence.
ACCESSION   BG355157
VERSION     BG355157.1   GI:13237143
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 256)
AUTHORS     Walbot,V.
TITLE       Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL      Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 947043  row: D  column: 12.
FEATURES             Location/Qualifiers
     source            1. .256
                       /organism="Zea mays"
                       /cultivar="B73"
                       /db_xref="taxon:4577"
                       /clone_lib="947 - 2 week shoot from Barkan lab"
                       /tissue_type="leaf and stem, including leaf base"
                       /dev_stage="2 week old seedling (3 leaves)"
                       /lab_host="XL1-Blue"
                       /note="Organ: shoot; Vector: Lambda ZAP (pBlueScript SK-);
                       Site_1: EcoRI; Site_2: XhoI; Directionally cloned using
                       Stratagene's UniZap XR cDNA cloning kit with the 5' end
                       at the EcoRI site. The library represents 8 x 10e5
                       independent recombinant phage. The plants were greenhouse
                       grown."
BASE COUNT      70 a      49 c      71 g      66 t
ORIGIN

```

```
Qy      92 agctgctatggtctgtgtcatgtacaccacctcggcacaaagcaggaaggagtggctaaa 151
        |||
Db       2 AGCTGCTATGGTCTGTGTCATGTACACCACCTCGGCACAAGCAGGAAGGAGTGGCTACAA 61

Qy     152 ctcgtagcaacctgatggaaggggtggatacaactctgttcccatcaacggcgggtggcag 211
        |||
Db      62 CTCGTACGAACCTGATGGAAGGAGTGGATACAACTCTGTTCCAATAAACGGCGGTGGCAG 121

Qy     212 ccctagctaggcgggtggatccgagcctgtatcagaaatcgaaataatataa 263
        |||
```

Db 122 CCCCTAGCCAGGCGGTG-----GAGCCTGTATCAGAAATCAAAAAAATATAA 168

Search completed: February 7, 2002, 08:20:48  
Job time: 18125 sec